Metaphor and the Brain: behavioral and psychophysiological research into literary metaphor processing

Hoorn, J.

1997

Metaphor
and the
Brain:

Behavioral and
Psychophysiological
Research into
Literary Metaphor
Processing

Johan F. Hoorn

Metaphor and the Brain: Behavioral and Psychophysiological Research into Literary Metaphor Processing

Dit onderzoek is verricht in het kader van het AiO-Netwerk Literatuurwetenschap, overgegaan in de Onderzoekschool Literatuurwetenschap: 'Empirisch onderzoek naar leesprocessen: de literaire metafoor psychofysiologisch benaderd.'

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Hoorn, J.F.

Metaphor and the Brain: Behavioral and Psychophysiological Research into Literary Metaphor Processing
Thesis Vrije Universiteit
-With ref.- With summary in Dutch.

ISBN 90-803717-1-8

Subject headings: Literature, Psychology, Psychophysiology, Metaphor

VRIJE UNIVERSITEIT

Metaphor and the Brain: Behavioral and Psychophysiological Research into Literary Metaphor Processing

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
prof.dr. E. Boeker,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der letteren
op maandag 9 juni 1997 om 15.45 uur
in het hoofdgebouw van de universiteit,
De Boelelaan 1105

door

Johannes Ferdinand Hoorn

geboren te Haarlem

Promotoren: prof.dr. E. Ibsch
prof.dr. A.F. Sanders

De kwaliteitsbeoordeling van dit proefschrift
door prof.dr. W.J.M. Levelt, verbonden aan de Katholieke Universiteit (het Max-Planck-Institut für Psycholinguistik), Nijmegen, wordt onder dankzegging vermeld.

Acknowledgements

First, I would like to thank prof.dr. Andries Sanders. If I have learnt anything during this study, it is from him. In my opinion, he clarified the consequences of metaphor theory better than the authors of these theories did themselves. He kept me from stumbling into about every pitfall empirical research entails. If there are any faults left, they are mine. He was the best that happened to me.

I would also like to express my gratitude to prof.dr. Elrud Ibsch for her open-mindedness. It is not at all obvious that a scholar of literature would support such an extreme empirical viewpoint as displayed in this study on metaphor. Without her role as fiery advocate, none of this would have been possible.

Furthermore, I enjoyed the cooperation with the group of Cognitive Psychology, in particular with drs. Floor van der Ham (mathematical support and theory modeling), dr. Dirk Knol (abundant statistical support), dr. Jaap Woestenburg (psychophysiological support), dr. Rob van der Lubbe (for helping me stepping into InstEP), and drs. Jan Willem Kleisen (the ERTS miracle). For ample discussions on the empirical approach to literature, I would like to thank my colleagues in the Department of General and Comparative Literature, particularly dr. Dick Schram, dr. Gerard Steen, dr. Charles Forceville, and dr. Paul Buurman. Prof.dr. W.J.M. Levelt, prof.dr. L.G.M. Noordman, dr. Gerard Steen and Shaike Shen Ph.D. are gratefully acknowledged for their valuable reviews of this thesis. David Hanauer Ph.D., Lindsay van den Burg and particularly Daniel Carroll M.A. all accounted for a substantial part of the English corrections, for which I thank them so much. Many of these people not only were my colleagues, but became close personal friends.

Without subjects, no tests; without tests, no science. Therefore, I apologize to the numerous students from the Language and Literature Departments of the Free University, and the Municipal Universities of Amsterdam, Utrecht and Leiden for torturing them in sometimes rather disagreeable experiments.

CONTENTS:

PREFACE

A two-way procedure of reading

In view of its interdisciplinary character, this book allows the reader to read it in two ways: Parallel or serial. Parallel reading is intended for scholars of literature and psychologists who are not interested in each other's field of study. Serial reading is for those who are interested in the empirical study of literature.

This book consists of seven Chapters, an introduction and a conclusion. Each Chapter addresses a specific issue of metaphor theory. Chapters 1 up to 3 have four Sections, which are fully theoretical, while Chapters 4 up to 7 have seven Sections, including tests and discussions. The Sections in each Chapter address the steps taken in the research cycle to investigate the issue of a Chapter. Each Section number has the following:

- .0 : Outline of a general problem in metaphor theory. (Questions asked about metaphor)
- .1 : Theoretical approaches of the problems under .0 in the study of literature
- .2 : Theoretical and experimental approaches of the problems under .0 in psychology
- .3 : Presentation of the theories used in the present study, based on .1 and .2 and their predictions
- .4 : Tests on the predictions in .3 and their results
- .5 : Discussion of the results for psychological theory
- .6 : Discussion of the results for the theory of literature

Traditional scholars of literature are advised to read the Sections .0, .1, .3, and .6 of each Chapter, parallel to psychologists with no interest in the theory of literature, who are advised to read the Sections .0, .2, .3, .4, and .5. Those who are interested in the empirical study of literature are advised to read all Chapters in serial order, that is, in order of appearance.

INTRODUCTION[*]

Metaphor and the brain

Organisms with a central nervous system have specialized areas in the brain that can detect features. These features can be lines, edges, light contrasts, sounds, or, even hands and faces in monkeys and probably in humans as well. Any two central nervous systems of the same species can communicate. Humans do this predominantly by means of language. Language starts by uttering sounds in response to the world, or to the concept of this world. This concept of the world is gradually established and refined as stable and coherent patterns of features. These patterns are made known to the brain as neuron-firing frequencies from perception cell to brain cell. According to this view, learning what an object is, is to distinguish it from another object through its features, evoking a unique pattern of firing frequencies, which target at the appropriate feature-detection cells. Thus, the concept is formed by unique frequency patterns, which repeatedly stimulate a combination of detection areas. In line with this idea, it could be that anything evoking a unique combination of firing frequencies is called an object. If the unique frequency pattern of an object matches the expected pattern of a concept, the object is judged as 'true' or 'real'. If it does not match the concept at all, it is judged as 'false' or 'unreal'. If the pattern of frequencies matches partly, as it does most of the time, it is judged as 'probable' or 'may be'. An evolving change of the concept passes through the latter possibility. A sudden change, or revolution, skips this, what may be called, 'metaphorical phase'.

Literal expressions move in the area of 'true' and 'real'. Journalism and science pretend to operate in this area by trying to make their statements fit the world as closely as possible. These statements need thorough research and ample discussion to make the concept fit the object. Metaphoric expressions move in the area of 'probable' or 'may be'. If a scientific statement is still debated, it is half a metaphor. 'Man is a machine' is a fine example. Metaphoric expressions are neither 'true' nor 'false'. Therefore, metaphors can be viewed as a continuum between 'highly probable' and 'highly improbable'. Scientific metaphors, such as process models of the brain, move in the area of 'probable' and 'true'. Literary metaphors, such as those found in poems and novels, move in the area of 'improbable' and 'false'. This is the reason the poet's claim is correct that he 'lies the truth'. In literature, metaphors need the artifice of the logical predicate [IMAGINE], to avoid it from being 'false' and 'unreal', this is, the area of the anomalies (cf. Levin 1977: 119). This does not mean that the [IMAGINE] predicate states 'this is a metaphor', but rather that metaphors need the conception of an imaginative world to be acceptable.

[*] Notes are on page 11.

Metaphor founds poetry

If poetry is not characterized by the predicate [IMAGINE], it could be taken literally, and therefore, easily rejected by the reader as pure nonsense. This explains why readers with little imagination do not like to read literature (this implication cannot be reversed). People who buy a novel that says 'novel' on the cover, buy a book with the predicate [IMAGINE] on the hardback. Any writer who wishes to violate the genre has to change this predicate.

Poetry often tends to be highly imaginative, and because a metaphor is neither 'true' nor 'false', it is an outstanding instrument for enhancing imaginativity and avoiding poetry from the claim of truth value. Since poets do not wish their poems to be rejected on these grounds, the area of rejection moves to beauty, imaginativity, or 'figurativeness'.

In the past, the study of literature has focused on metaphor, because it is seen as a founding feature of literature. This focus is so strong that metaphor became the pre-eminent instrument for describing and analyzing literature. Traditional approaches, such as Russian Formalism, Prague Structuralism, Semiotics and Reception Aesthetics, never took much trouble about validating their theories. Consequently, their statements are neither 'true' nor 'false', which render these theories highly metaphorical, closer to art than to science. Since poetry and metaphor are described by these highly metaphoric meta-languages, the paradox occurs that the unvalidated theoretical statements are neither 'true' nor 'false'. On the other hand, they are not poetry either, because they lack the [IMAGINE] predicate. Thus, the study of literature so far is a highly philosophical one.

However, through the ages, various groups of metaphor theories have emerged, based on different assumptions about the textual function of metaphor. One issue that has always remained implicit in the theory of literature, is that the textual function of a metaphor (connections between metaphors and other textual elements) depends on cognitive processes involved in understanding the metaphor (how the metaphor is interpreted). Reading is cognitive processing. Hence, any theory about the textual function of the metaphor is implicitly a cognitive theory about processing metaphors.

In this study, the three most important theoretical approaches to metaphor processing are analyzed and formalized in three competitive models: Comparison theory, anomaly theory and interaction theory. Chapters 1, 2, and 3 will discuss these models.

Metaphor as an understanding device

The best way to get acquainted with a new concept is through an old one. If someone wants to learn about elementary particle physics and does not know what 'quarks' are, a basic understanding of it can be formed by presenting 'quarks as building blocks' (Huang 1992: 25). Building blocks actually are parts of houses and offices. However, used as a metaphor, they may explain something about the function of quarks in an atom.[1] To give a further example of metaphor as a means of explanation, a yet unknown phenomenon of lightning flashes above the clouds was described by E. Westcott of the Alaska University as giant 'roots' or 'jelly-fishes' (October, 1993). Westcott reported that the flashes were brightest at the top, about sixty kilometers above the ground - where the head of the jelly-fish was sited - with its tentacles hanging down. 'The lightning flash is a jelly-fish' is a nice example of utilizing metaphor as an understanding device. The topic is yet unknown to the audience, because the lightning flash is not an ordinary one. It has uncommon features, which are quite difficult to describe literally. Instead, the researcher uses a concept that is already known in detail to the audience: The jelly-fish. Features such as 'round at the top' (the head), 'long flexible forms hanging down' (tentacles) are activated by the concept of a jelly-fish, so that a first idea about the new lightning type can be formed.

Metaphors can also help science to form a new concept about an entity it is in search of, but which has not yet been found (for example, 'the wonder tissue' in the brain). Chemists speak of 'the magic island in a sea of instability', when referring to superheavy nuclei of atoms which can be manufactured artificially. These elements are heavier than uranium and hard to stabilize.

> The three-dimensional map (...) shows very clearly how the stability of atomic nuclei varies as the numbers of protons and neutrons increase. There is a "mountainous peninsula" of stability running diagonally across the map. In the lower part of the map, stable nuclei have roughly equal numbers of protons and neutrons. Further up, they have more neutrons. The mountain peaks are the so-called magic numbers - numbers of protons and neutrons producing the extra stable closed shells. Surrounding the peninsula is a sea of instability. At the far end, at atomic number 114 is the "island of stability" (...) (Loveland & Seaborg 1991).

The first three metaphors in this quote are based on the knowledge that chemistry does have about atomic nuclei. The first says that 'the range of stable nuclei is a mountainous peninsula'. The second, 'mountain peaks are magic numbers' actually is constructed from two metaphors, where 'magic numbers' is a metaphor for the number of protons and neutrons to make an extra stable nucleus, and 'mountain peaks' is a metaphor for 'magic numbers'. The elements that are not stable are called the 'sea of instability' in which the 'mountainous peninsula' lies. 'The sea of instability' is a mixture of experimental knowledge and speculation about the elements that already proved to be unstable, and those that are expected to be, but are not produced yet. Together, the three metaphors form the overall metaphor, that the variety of proton and neutron combinations can be represented as a geographical 'map'. The fourth metaphor on this 'map' is the most interesting one. It is a concept legitimized by the map-metaphor, yet without having a real reference. The 'island of stability' or 'magic island' is a belief about those extra heavy elements which are expected to be stable, but which are not proven to be as yet. Features from the already known range of stable nuclei (the 'mountainous peninsula') are expected to be found for the 'magic island'. It should have a mountain peak, similar to the ideal number of protons and neutrons in the stable nuclei. And although the magic island lingers in the realm of the unknown, the chemists 'have been trying to make a leap to this island, hoping to make a new range of elements'. When the attempts to make this 'leap' fail, they are 'crashing into the sea of instability' (Loveland & Seaborg 1991).

Thus, metaphors may be used to launch a hypothesis about an unknown world, conceptualizing the features of what should be found. These concepts could turn out to be a unicorn. Something which is sought, but which can never be found, because the concept has no reference to a 'real' object. The object of reference needs the predicate [IMAGINE] to justify its existence and thus the concept enters the realm of metaphor. Metaphors can become literal statements after testing, so that a truth value can be associated with them.

Cognitive psychologists are becoming increasingly aware that metaphors are a way of understanding the world, as alternative to 'true' statements. Also, cognitive psychologists are becoming aware that language is highly metaphoric, because the internal concept never fully describes the object. Even scientific languages describe their objects in terms of mathematical metaphors.

Cognitive psychology has been concerned with the empirical study of language, and has a well established experimental tradition. One goal of psycholinguistics is to investigate the basic elements of understanding language, before entering into the complicated matters of full text processing. This means that their studies mostly concern issues such as word frequencies, phonological encoding, lexical ambiguity and, recently, word priming in short sentences. Psychologists lay the bricks, before building the house. To investigate the fundamentals of language processing, artificial stimuli are used, which manipulate the linguistic aspects of interest. This is a perfectly legitimate way of achieving one's scientific goals. However, the problem with studying literature is that the object of study is always a 'natural' text; in other words, a text not written for experimental purposes. One way of dealing with the problem is to cease studying literature until more is known about reading. The other way is to use natural texts in contrast to artificial ones and to investigate their different effects. This also means that literary texts which are used in experiments must be submitted to a series of tests, before a set of useful replications can be carried out.

Aims and definitions

> (...) the main problem is to know what happens when actual recipients attribute meaning to texts which they conceive of as literary texts. This question must be answered with as much precision as possible before any claim can be made as to the supposed function and effect of literature. (Ibsch 1991: 4)

The aim of this research is not to describe the social function or cultural role of metaphors, but to study metaphor processing in different textual conditions. In metaphor theory, three directions are much in evidence: *Comparison theory*, *anomaly theory*, and *interaction theory*.[2] In Chapter 1, 2, and 3, these three theories are treated as three competing models of metaphor processing. The predictions deduced from the theories are tested in Chapter 5 up to 7, in order to evaluate their adequacy for processing (literary) metaphors.

The theoretical study is based on the work of some major scholars of literature, and on more detailed psychological studies into metaphor processing. The road of heuristic study is left explicitly, subscribing Jakobson, Fant & Halle's statement on the study of phonetics that language research:

> (...) requires experimental verification and further elaboration. The nature of these problems calls for coordinated research by linguistics, psychologists, experts in the physiology of speech and hearing, physicists, communications and electronics engineers, mathematicians, students in symbolic logic and semiotics, and neurologists dealing with language disturbances, as well as the investigators of the poetic use of speech sounds. (Jakobson, Fant & Halle 1963: v)

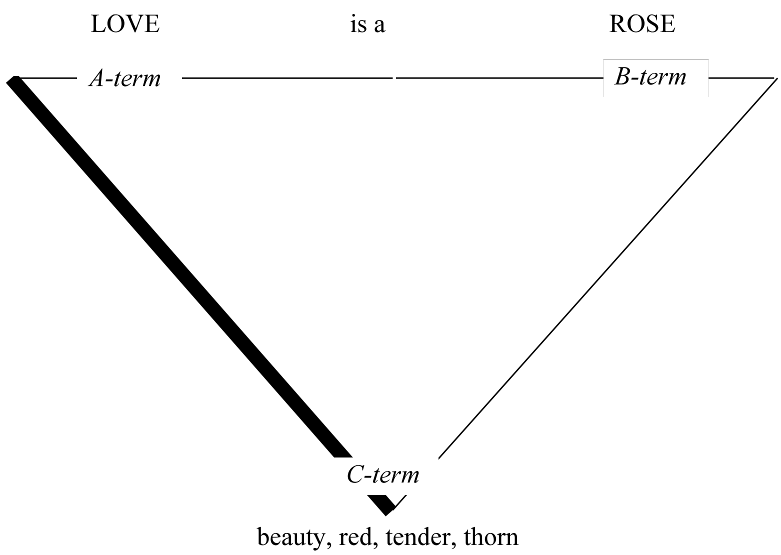
Therefore, the experimental research of the present study is based on psychological investigations, such as set theoretical approaches to similarity, reaction time (RT) studies into metaphor processing, and event related brain potential (ERP) studies into semantic anomalies.

In the next Sections, certain concepts are defined: Literal, metaphoric, and anomalous expressions, features, context and priming. These definitions are formulated without intentionally biasing one of the metaphor theories.

Definition 1: A-, B-, and C-term

The literal, metaphoric and anomalous expressions in this research, consist of three parts: An *A-term*, a *B-term*, and a *C-term*. The *A-* and *B-term* are always explicitly present in the text, and are connected by the auxiliary 'to be' (*A* is *B).* The *C-term* is the great unknown. Most theories claim that it is the basis on which the *A-* and *B-term* are compared (traditionally, the 'tertium comparationis'). According to certain theories, the *C-term* is based on or even identical to the set of features that is shared by *A-* and *B-term*. What is considered the *C-term* strongly depends on interpretation. In a metaphor, such as 'love is a rose', the *A-term* is 'love'. The thing with which love is compared is the *B-term*, 'rose'. The *C-term* may be something like 'beautiful, red, tender, and thorny' (see Figure below for this triangle relation).

The triangle relation among the elements of a metaphor. The *A-term* is the main issue of the sentence, not necessarily its grammatical subject. The *B-term* is the image in which the *A-term* is perceived. The *C-term* is the meaning of the sentence, i.e. the solution to the metaphor. Which term is seen as *A-* or *B-term*, or what the *C-term* might be, is a matter of reader interpretation.



In all cases, the *A-term* is the central issue of the expression, and the *B-term* is the frame of reference, while the *C-term* is the basis on which the comparison between *A-* and *B-term* is legitimatized.[3] Notice that *A-* and *B-term* do not necessarily coincide with grammatical subject and nomalized predicate (although often, they do). In other words, a fixed order of *A-* and *B-term* may not be assumed a priori. Therefore, reader judgements should decide which word in an expression is the *A-* or *B-term* (cf. Table 4.1, Chapter 4).

Definition 2: Literal, metaphoric, and anomalous expressions

Three expression types will be investigated: Literals, metaphors, and anomalies. Merely tentative definitions are in place here, because the very characteristics upon which the expression types are identified is the object under study.

A literal expression usually matches an instance with an appropriate category ('love is an emotion'), and depends on a truth value, whereas a metaphor usually matches inappropriate instances and/or categories, and is independent of truth value. An anomaly may also match inappropriate instances and categories. However, it is not a metaphor, but is 'beyond imagination'.

For most philosophers and literary theorists, 'literal' expressions do not exist. Since truth cannot be obtained, they believe that 'all is metaphoric'. In this study, 'literal' is defined as language that is conventionalized and lexicalized as describing the 'real world'. In the experiments, however, subjects will judge for themselves what is understood as describing the 'real world'. In other words, 'literal' is defined subject-dependently, and not in absolute terms.

As in psychology, the theory of literature distinguishes between idiomatic and novel metaphors. Idiomatic metaphors are commonplaces, such as 'love is a rose', 'this boy is as obstinate as a mule' or 'she fought like a lion'.[4] The novel metaphor is newly created and has not yet come to stay (e.g., 'poetry is a matador and metaphor its bull'). This study is restricted to the novel metaphor in literary texts. A metaphor is emphatically not defined as a logical proposition, because a logical proposition presupposes a fixed division between subject and predicate and claims a truth value.

'A rose is a blip' may be an example of an anomaly, and is the mismatch of a word with contextual expectations, which renders the expression highly nonsensical. This mismatch can be formed in many ways, one of which is a between-category mismatch. In the case of 'a rose is a blip', the term 'rose' originates from botany and 'blip' from radar technique, two categories which are unlikely to be connected. However, a within-category mismatch also may lead to anomalous language use. In 'a rose is a tulip', the category of both exemplars is 'flower', and putting on a par two exemplars of the same category, apparently results in a semantic violation. An explanation may be that the features which distinguish both exemplars within the common category are erroneously equalized by the suggestion of identity, thus making the distinction fail (Sections 4.2, 4.5 and 4.6 will further discuss the relation between category verification and semantic deviations). Whether or not an expression is judged literal, metaphoric or anomalous, strongly depends on the individual's own level of tolerance, so that in this respect, decisions should be left to subject groups.

Definition 3: Features

Every word has meaning connections with other words. All the meanings, words, and associations which a word evokes are called 'the features' of this word. These features may be on a formal level (syntax, morphology, spelling, phonology) as well as on a meaning level (symbols, semantic connections). The combined features of a word make up the feature set of that word. The feature set of a word, thus can be seen as a long list of words summed up by readers in response to a word.[5]

The feature set of a word has common and personal components. The common components refer to the meaning of the word according to the dictionary, the standard language shared by most members of a speech community. These are the high-frequent associations, or high-frequent features in the feature list of the word.

The personal components are the idiosyncratic meanings and associations that an individual member of the speech community attaches to a word. These are the low-frequent associations, or low-frequent features in the feature list of the word. Thus, in a metaphor, the features of the *A-* and *B-term* may be any (formal or semantic) association activated (or created) by any member of a speech community.

Definition 4: Context = text

Context is usually described as all the social and cultural conventions in which a text functions. In this study, the word 'context' is not used in that way. Context will be used for the notion of text; one or more sentences, or even one word, which spatially and/or semantically neighbours another word. Context in this sense is understood as linguistic context. Therefore, in this study, the notions 'linguistic context', 'context' or 'text' all mean the same. An experimental trial may occur under three conditions. If the trial consists of an isolated *A-* or *B-term*, the condition is called a *single term* condition. If an experimental trial consists of a single expression, the condition is called an *expression* condition. If an expression is presented as part of a larger text, this is called a *context* condition.

Definition 5: Priming

Priming is an important function of context. Priming refers to the whole of textual elements that favors one interpretation of the *B-term* of an expression and thus, favors one interpretation of the whole expression (e.g., a literal or a figurative meaning). The priming effect of the original literary text is unknown. In *expression* and *context* conditions, targets are always formed by the *B-term*, while the *A-term* and all other text form the prime.

Notes:

1. M. Gell-Mann - one of the most important post-war particle physicists - claimed to have taken the term 'quark' from James Joyce's novel *Finnegans Wake* (Nambu 1985: 104-105).

2. Mooij (1976) offers a review of the philosophical literature on the three prevalent schools. Verbrugge & McCarrell (1977) and Tourangeau & Sternberg (1982) offer a review of the psychological literature.

3. In metaphor theory, the terminology for the *A*-, *B*-, and *C-term* is rather confused. The *A-term* is, among others, also called 'tenor' (Richards 1965: 96), 'topic' (e.g., Verbrugge & McCarrell 1977), or 'focus' (M. Black 1980: 28). The *B-term* is often called 'vehicle' (Richards 1965: 96), or 'frame' (M. Black 1980: 28). The *C-term* is also referred to as 'ground' (e.g., Tourangeau & Sternberg 1982) or 'tertium comparationis' (e.g., Verbrugge & McCarrell 1977; Wolff 1977: 54).

4. Henle (1966: 187) states that "metaphors of this type tend to vanish, not in the sense that they are no longer used, but in the sense that they become literal, so that today no one would think of saying that 'plastron of a turtle' or 'hood of a car' were metaphors". Indeed, there is psychological evidence that idiomatic metaphors like 'bury the hatchet' are understood as easily as a literal expression. Activation of the original figurative meaning of the idiomatic metaphor slows down the understanding (Ortony, Schallert, Reynolds & Antos 1978; Estill & Kemper 1982).

5. In the relevant literature, features are also called 'properties' (e.g., Beardsley 1982: 271) or attributes (e.g., Gentner 1988).

CHAPTER 1: THE COMPARISON MODEL*

1.0 **Comparison theory**

Metaphor theories attempt to explain how a metaphor is understood. This is usually done by proposing a set of processes that are supposed to underly the interpretation of a metaphor. Some theories state that the 'false', 'figurative' *B-term* is nothing but a substitute for the literal term. These *substitution* theories claim that a metaphor does not express more or something different than a literal expression about the same issue.

However, substituting terms without changing the meaning only holds if the meaning of the substituting term is identical to the substituted term. In 'the boy is a weasel', 'weasel' can be replaced by 'coward', although 'weasel' is not identical to 'coward'. Usually, a substitute is a *similar* term, not an identical one. A substitute can be identical, but not necessarily. Unless they have exactly the same features, substitutes do not express identical meanings.

Comparison theory does not share the idea that a metaphor is merely another form of literal speech. The comparison view usually sees a metaphor as an implicit simile. Similes are comparisons, marked by an explicit 'as' or 'like' construction.[1] Comparison theory claims, that in order to understand a metaphor, the two terms of a metaphor are compared to find common features. Sections 1.1 and 1.2 will focus on the core assumptions of the comparison theory. These assumptions will be formalized in Section 1.3 in a comparison model of metaphor understanding, in which the perceived similarity of a metaphor is an assigned value based on frequencies of shared features.

1.1 **Comparison theory in literature**

The first traces in Western philosophy of a comparison view of metaphor understanding are found in Aristotle's *Poetics* and *Rhetoric*. In *Poetics*, Aristotle spent just a few lines on metaphors, but they summarize the comparison view quite accurately:

> Metaphor is the transference of a term from one thing to another: whether from genus to species, species to genus, species to species, or by analogy. (Halliwell 1987: ch. 21, 55)
>
> (...) the successful use of metaphor entails the perception of similarities. (Halliwell 1987: ch. 22, 57)

---

* Notes are on pages 24-25.

In the above quotations, several mechanisms are offered to explain metaphor understanding: 'Transference of a term', 'analogy', and 'perception of similarities'. First, some confusing terminology should be clarified. In the quotation, 'term' does not mean *A*- or *B-term*, but 'feature'. Thus, the mechanism is 'feature transfer'. By 'similarities' is meant 'shared features', which may cause the perception of similarity or semantic relatedness. Thus, the mechanism is 'finding the shared set'.

'Feature transfer' means that the *B-term* adds a feature to the *A-term* that was not pre-existent in the *A-term* set. In 'the boy is a rat', the features 'nasty' and 'sneaky' are taken from the *B-term* and added to the *A-term* set, so that the *A-term* ('boy') is also qualified by the features 'nasty' and 'sneaky'.

'Shared set' means that the feature sets of *A*- and *B-term* are compared to find common features between 'boy' and 'rat'. In this approach, 'nasty' and 'sneaky' are already present in the *A-term* set, although this may be on a low activation level; as latent or 'possible' features. Sometimes, finding the shared set is understood as finding the common category. In that case, 'boy' and 'rat' both belong to the category of 'treacherous creatures'.

The notion of analogy is not pursued is this Chapter, because it is the principal mechanism proposed by the interaction theory (Chapter 3). However, Section 1.3 gives a short account of the idea.

All these mechanisms should lead to an increase in the semantic relatedness (hence similarity) between the *A*- and *B-term*. In the comparison theory, the perception of similarity is seen as an immediate occurrence to, rather than a *creation* of the mind. An example of this view is found in Middle Eastern philosophy on metaphor. Influenced by the work of Aristotle, a major philosopher on poetic syllogism,[2] Farabi (870-950), stated:

> For [poets] compare A to B to C, because there exists between A and B a close, fitting, and familiar likeness, and there exists as well between B and C a close, fitting, and familiar likeness. So they gradually unfold their words until they make occur to the listeners' minds a relation of likeness between A and C, although originally they were remote. (D.L. Black 1990: 214-215)

The quotation above states that there is a likeness between *A* and *B* and between *B* and *C*. This likeness (or similarity) is close, fitting and familiar and it already 'exists'. In other words, the similarity between two terms (*A-B*, *B-C*) is not created, but is fixed and already present. Since one way or the other, similarity is supposed to be related to the feature sets, the relevant aspects of these feature sets should also be fixed and pre-existent. For instance, if similarity is dependent of the shared features, then these features do not need to be created by the listener or reader; they just need to be found. If shared features already exist and just need to be found, this means that they are always the same in a metaphor, independent of context. Consequently, if the shared set remains the same under all conditions, it follows that all terms activate (at least partly) fixed feature sets, which do not change as a function of context.

In more modern approaches to metaphor understanding, remnants of the comparison view are still met. Often, the modern approaches are mixtures of comparison theory, and notions of anomaly theory or interaction theory. Mooij (1976), in arguing against the comparison view, reluctantly admitted the merits of comparison theory in the course of his discussion on metaphor:

> (...) metaphors are often said to be based on similarity and analogy. However, I shall not make use of the last mentioned assumption. (Mooij 1976: 22)

> (...) the idea that metaphors are based on similarity or analogy entails the disadvantage that it seems to favour a certain view on metaphor, viz. the view that metaphor is a shortened comparison. Strictly speaking, one can accept similarity and analogy as a basis without attaching the said consequence (Mooij 1976: 23).

> It is easy to understand, (...), that similarity and analogy have often been considered the basis of metaphor. It is undeniable that similarity or analogy is hidden in many metaphors. (Mooij 1976: 23)

In a pure form, comparison theory is hardly found in modern metaphor theory. The more modern examples are discussed in Chapter 2 (on anomaly theory) and Chapter 3 (on interaction theory). In these Chapters, however, anomaly theory and interaction theory will be discussed in the light of comparison theory to point out the close relationship with this older view.

1.2 **Comparison theory in psychology**

As mentioned in 1.1, the comparison view can be traced back to the Aristotelian approach as described in *Poetics* and *Rhetoric* (Rhys Roberts 1966; Halliwell 1987). The comparison view sees a metaphor as an implicit simile, which is a comparison with 'as' or 'like'. The two terms of the metaphor, the *A*- and *B-term*, have a certain 'similarity', which is supposed to be the solution to the metaphor, its *C-term*. When a *C-term* is found, the metaphor is understood. Comparison theory suggests the following mechanisms that could underlie the forming of the *C-term*.

'Feature transfer'. Comparison theory claims that sometimes features are transferred from one term to another. If a term of a metaphor is not very well known to the reader (most of the time, this is the *A-term*), features of the other term (usually the *B-term*) are transferred to the first to improve understanding. In 'he is a weasel', the feature 'cowardice' is assumed to be transferred from 'weasel' to the unknown 'he'.

'Shared set'. Another idea is that the terms activate two fixed feature sets. Between those sets, identical features must be found to form the *C-term*. Experimental studies on similarity perception often opt for the 'shared set' approach (e.g., Tversky 1977; Malgady & Johnson 1980). According to the comparison view, shared features form a fixed set. The feature sets activated by the *A-* and *B-term* are fixed as well. If the activated features and the shared set are always the same - regardless of any other factor - context should not influence the number and kind of features that are activated or overlapping. 'Finding the shared set' is sometimes exchanged by 'finding the common category', in which case 'he' and 'weasel' both should be a member of the category of 'cowards'. 'Metaphor by analogy' is the basic principle of the interaction theory (Chapter 3), and will be shortly discussed in Section 1.3.

Witness to his ideas how the *B-term* is associated with the *A-term*, Skinner (1957) may be seen as a representative of the comparison view on metaphor processing in psychology.[3] By emphasizing the importance of similar properties (features) between terms, he took a comparison point of view:

> We do not need to say that the speaker "discovers a similarity and expresses it by transferring a response." The response simply occurs because of the similarity (Skinner 1957: 92).

The similarity to which Skinner referred is a 'similar stimulation', produced by two objects (the object the *A-term* refers to and the object the *B-term* refers to). This similar stimulation leads to the response that the second (the *B-term* object) 'is like' the first. The point is illustrated by two examples:

> In *Juliet is* [like] *the sun* we must explain the appearance of the response *sun* when no sun is actually present. We do so by noting that Juliet and the sun have common properties, at least in their effect upon the speaker. (Skinner 1957: 93)

> If we first acquire the response *leg* in connection with animals and extend it to legs of tables and chairs on the basis of geometrical and functional similarities, the properties common to all these cases acquire control of the response and are subsequently respected by the community. (Skinner 1957: 94-95)

In other words, Skinner sees the *B-term* as an almost necessary response to the *A-term*, because of its similar properties (or features). If this is true, context cannot have a dramatic effect on the activated feature sets, because then, the *B-term* is not automatically associated with the *A-term* any more.

According to Skinner, processing literal and metaphoric expressions is basically the same. It is represented by the 'three-term relation' stimulus-response-reinforcement. Skinner did not see metaphor processing as a kind of analogical reasoning. Special features in the *A-term* decide, whether the associated *B-term* is literal or metaphoric:

> A second type of extension takes place because of the control exercised by properties of the stimulus which, (...), do not enter into the contingency respected by the verbal community. This is the familiar process of metaphor. Traditional accounts, from Aristotle on, have generally assumed that, like generic extension, metaphor is a special achievement requiring a special faculty of analogical thinking. But the basic process is again adequately represented by our three-term relation; the only difference between metaphorical and generic extension is in the kind of property which gains control of the response. (Skinner 1957: 92)

This would lead to the consequence that not every two words can be interpreted as a literal or metaphoric expression, but that some words are better suited to become a metaphor, and others to become a literal statement. Again, if metaphoricity is demanded from the isolated *A-term*, contexts will not have a great influence on which *A-term* features 'cause' the *B-term*. Some words would lead to a metaphor and others to a literal expression. Unfortunately, Skinner was not explicit on the context matter, and did not explain which kind of feature gains control of the response.

1.3 A formalized model of the comparison theory

In summary, the core assumptions of comparison theory are (cf. Section 1.1 and 1.2):

1. A metaphor is not a literal statement
2. The surface form of a metaphor consists of two terms
3. A metaphor is an implicit comparison *A is like B*
4. Transformation into an explicit comparison is a one-stage process
5. Each term evokes a fixed feature set
6. These terms share features, which form the *C-term*
7. The shared features form a fixed set
8. Given this, there should be a criterion for the number of shared features that makes an expression literal, metaphoric or anomalous
9. There is no effect of context on the (shared) feature sets

Regarding point 4, the comparison theory is not explicit about the decisive criterion to transform an expression into the form '*A* is like *B*'. If someone reads an implicit comparison, it may also be a literal expression. Therefore, point 4 induces that some criterion is met to signal the encounter of an implicit comparison, and not of a literal expression. The problem may be solved by starting the process with the activation of the two feature sets, the number of shared features between which meets a certain criterion to decide whether an expression is literal or metaphoric.

Theories of metaphor are - sometimes implicitly - based on the contrast with a theory of literal expressions. Literal expressions are usually seen as descriptors of the 'real' world. Words which are not marked by an [IMAGINE] predicate - such as 'man', 'woman' and 'child' - are seen as literal. Only if the [IMAGINE] predicate is active, 'man', 'woman' and 'child' could turn out to be a centaur, a mermaid, and a will-o'-the-wisp. A somewhat more complex descriptor, such as a definition, uses at least two descriptive words, for example, 'a ring is a circle'.

According to the comparison view, a metaphor has a certain similarity between the two descriptive words, but this similarity is insufficient to make the expression literal. Thus, the insufficient literal expression needs the [IMAGINE] predicate to save it from being erroneous or anomalous. Comparison theory solves this problem by the introduction of a nondefining device. The 'as if' construction makes the expression a comparison, rather than a definition. Owing to this, the insufficient literal expression has become a simile, and is acceptable again. Despite the smaller number of shared features, this is not a reason to read the expression as an anomaly, rather to read it as a metaphor.

In the comparison theory, the similarity between the terms of a literal expression is high. Therefore, the feature sets of a literal expression should be almost identical. Its extreme is a tautology. If not in its extreme, an expression needs a sufficient number of shared features to be literal.[4]

The shared set of literal expressions should be large. A metaphor has a certain similarity between the two terms, but this similarity is insufficient to make the expression literal. Despite the smaller number of shared features, this is not a reason to read the expression as an anomaly. An anomaly shows the least similarity, so that the shared set should be smallest.

Figure 1.0 shows a Venn diagram representation of the comparison theory for a two sets overlap of fixed features in literal, metaphoric and anomalous expressions. Figure 1.0 represents the *A-term* feature sets as smaller sets than those of the *B-terms*. It may be argued that the *A-term* is the focus of the comparison; the term of which something new is said. It may be that the *A-term* is less familiar, because it is clarified by the *B-term*, which functions as the frame of reference. Taking into consideration that the more one knows about an object, the more one can say about it (cf. Raaijmakers 1984: 177), familiarity may be correlated with the number of features that can be summed up. Thus, it would follow that the *B-term* evokes more features than the *A-term*. If the results are contrary to expectations, then, of course, the size of the feature sets are different from that suggested in Figure 1.0.

Figure 1.0: Comparison model for a two sets feature overlap in literal, metaphoric and anomalous expressions.

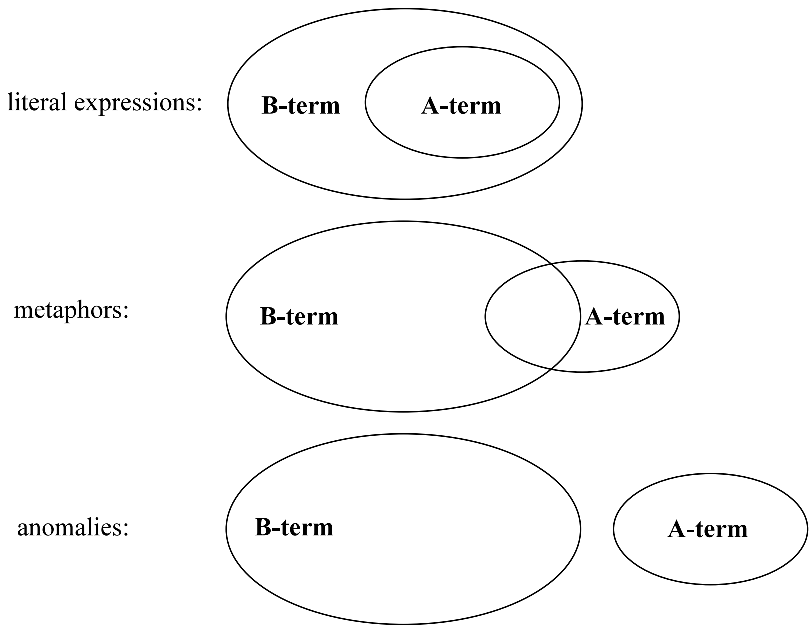


Figure 1.1 shows a flow diagram for processing literal, metaphoric and anomalous expressions, conform the eight assumptions of comparison the-

ory. The model consists of three phases: Encoding, comparing and responding.

In the *encoding phase*, the expression is perceived. The *A-term* is read, followed by the *B-term*. The *A-term* activates a feature set. The feature set of the *A-term* is called X. The size of the feature set X is designated by $\#X$ (the cardinal number of X). The *A-term* activates as many features as possible, which are indexed by n. Every time a new feature is found, the value of n is augmented by 1: $n = n + 1$. If no further features are activated, the activation of the *B-term* features starts.

The *B-term* feature set is called Y, and its feature set size is called $\#Y$. Again, as many features as possible are activated by the *B-term*, and every new feature augments $\#Y$ by 1: $m = m + 1$. Both loops for the feature activation of *A-term* and *B-term* express no more than 'get the first feature, get the second, get the third, etc., until there is no feature left' (end of file).

The encoding phase is represented by a fixed order serial process, as if the *A-term* feature set must be filled completely, before the *B-term* set is activated. Of course, it may be that the *B-term* set is activated in parallel with the *A-term* set. While the *A-term* is activating features, the *B-term* is read and activates its features synchronously.

When the feature sets X and Y have been activated, the *comparison phase* is entered. The search for shared features is performed by comparing each feature x in the *A-term* set X with each feature y in *B-term* set Y, to determine whether they match. As long as no common features are found, the shared set $X \cap Y$ (S, for short) contains no features. Thus, the shared set size $\#S$ is zero: $\#S = 0$. To compare the features of X and Y, the features x receive the index i and the features y the index j. The first feature x ($i = 1$) is compared with the first feature y ($j = 1$). The features x_i and y_j are compared: $x_i = y_j$?. If x_i and y_j match, one shared feature is found: $\#S = \#S + 1$. If they do not match, $\#S$ is not augmented. The search for shared features is also represented as a serial process. However, it is possible that all the loops of the model occur (partly) in parallel. For the train of thought, this is not important.

Next, the second feature of the *B-term* set is compared with the first feature of the *A-term* set. Therefore, the decision must be made if there are any features left in the *B-term* set Y: $j < m$?. If there is a feature left, $j = j + 1$, this *B-term* feature is compared with the first *A-term* feature. If they match, the second shared feature is found. This loop continues, until there are no *B-term* features left in set Y: $j = m$. If there is more than one feature in the *A-term* set X, the same procedure starts all over again for all *B-term* features, until there are no *A-term* features left to compare.

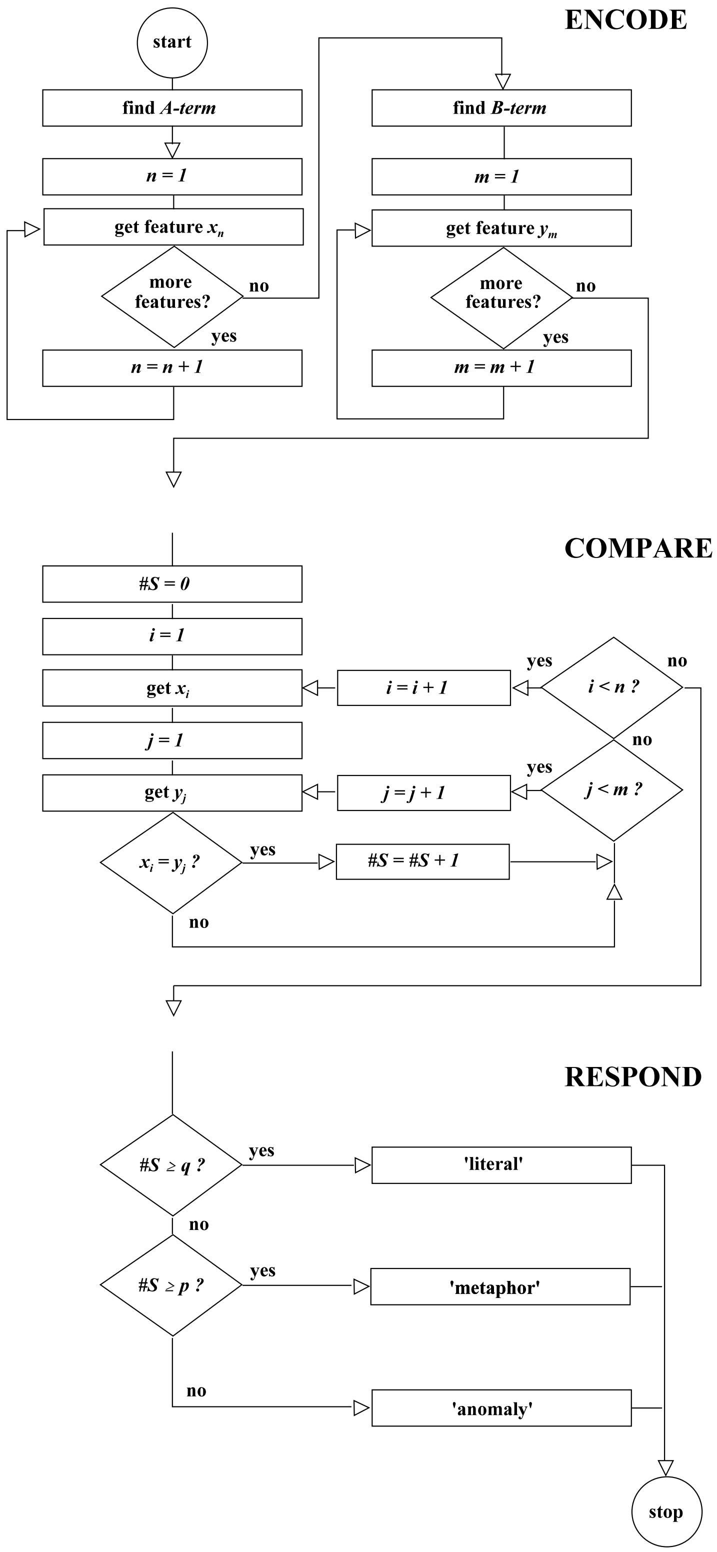
After the completion of the comparison phase, the value of $\#S$ is the size of $X \cap Y$ or the number of common *A-* and *B-term* features. According to the comparison view, $X \cap Y$ is the *C-term* of the metaphor, now expressed by a value which should correspond to the height of the perceived similarity between the terms. If there are no more *A*- and *B-term* features left to be compared, the response phase is entered.

In the *response phase*, the number of shared features is checked against a certain criterion that determines whether the expression is literal, metaphoric or anomalous. The criterion q is an internal 'threshold' value, corresponding to the Venn diagram for literal expressions in Figure 1.0. Criterion q demands a large feature overlap, so that the value of $\#S$ must equal or excede q, $\#S \geq q$?, to evoke the response 'literal'.

The criterion p corresponds to the Venn diagram for metaphors in Figure 1.0. Criterion p is also a 'threshold' value, but a lower one than criterion q. Criterion p demands a partial overlap of features, so that the value of $\#S$ must equal or excede p, $\#S \geq p$?, to evoke the response 'metaphor'.

If $\#S$ is smaller than p, the overlap of *A*- and *B-term* features is insufficient, and the expression becomes an 'anomaly'. The value of the criteria q and p is, of course, an empirical question. Since more criteria have to be checked, metaphors and anomalies are processed slower than literals. This ranking should make itself manifest in the respective response times.

Figure 1.1: Serial process comparison model for metaphor understanding (fixed order).



In summary, the formal representation of the serial process comparison model looks like this:

1. encode *A-term*: $X = \{x_1.....x_n\}$
2. encode *B-term*: $Y = \{y_1.....y_m\}$
3. compute $S = X \cap Y$
   if $\#S \geq q$ respond 'literal'
   else if $\#S \geq p$ respond 'metaphor'
   else respond 'anomaly'
4. stop

   q = number of elements needed in $X \cap Y$ to respond 'literal'
   p = number of elements needed in $X \cap Y$ to respond 'metaphor'

The model can be refined by using weighted features. This means that some features of a term are more important (e.g., more prototypical or higher frequent) than others. The weighing of the features, thus could lead to a different value of $\#S$. A weight may be assigned to a feature by a scale value, or the frequency of occurrence.

Hence, if u is a feature in X or Y or in the shared set S, $w_X(u)$ is the weight w of feature u in set X. Likewise, $w_Y(u)$ is the weight w of feature u in set Y. Therefore, the size (W) of the weighted feature set X may be described as the sum of all weights w of all features u in X:

$$W_X = \sum_{u \,\varepsilon\, X} w_X(u) \,,$$

while the size of the weighted feature set Y may be described as:

$$W_Y = \sum_{u \,\varepsilon\, Y} w_Y(u) \,.$$

Consequently, the weight w of feature u in S is a function g of the weight of u in X and in Y: $w_S\,(u) = g(w_X\,(u)\,,\, w_Y\,(u))$. This means that the shared set S may contain features that are differently weighted in X and in Y. For instance, the feature 'red' may be mentioned once for 'love' and twice for 'rose'. How the different weights of shared features should be treated arithmetically (summated, multiplied or divided) - in other words, how function g should be conceived of - will be discussed in Chapter 5.

It follows that the shared set size $\#S$, now can be rewritten as the weighted set size (W) of all shared features u, which is the sum of all weights $w_S\,(u)$ of every feature u in S:

$$W_S = \sum_{u \,\varepsilon\, S} w_S(u) \,.$$

The model in Figure 1.1 is designed as a fixed order process between feature activation and comparison. However, the weighing could change the rank order for the comparison of features, so that, e.g., high-salient features are compared before low-salient ones. As a consequence of this preferred order of comparison, the search does not need to be exhaustive to reach criterion *q* or *p*. In that case, these criteria should be checked first, before another comparison loop starts.

Another point is that the model suggests a time delay between checking *q* and *p*, yielding slower decision times for metaphors than for literals. However, it could be argued that *q* and *p* are checked at the same time, so that RT differences between literals, metaphors and anomalies will fail. The model should then be revised such, that the arrow from the comparison phase splits in two, and points at the decision diamonds for *q* and *p* in parallel.

Although Figure 1.0 suggests larger shared sets for metaphors than for anomalies, Figure 1.1 does not translate this into an RT difference. Both responses ('metaphor', 'anomaly') follow from the same decision diamond. This may be regarded as a shortcoming of the model - and so it is. It would be desirable also to differentiate metaphors from anomalies in response times (anomalies slower than metaphors). However, as the empirical findings show (Chapter 6), this question will turn out to be purely academic.

From the model in Figure 1.1, predictions can be made about the processing of metaphors, according to the comparison view. These predictions compete with those in Section 2.3 (anomaly model) and Section 3.3 (interaction model).

The predictions derived from the comparison view are:

1. **A metaphor activates two feature sets**
2. **These two feature sets are always fixed**
3. **A literal expression has more shared features than a metaphor and a metaphor has more shared features than an anomaly**
4. **The set of shared features is fixed (and thus, can be predicted from the sets of single terms)**
5. **The response time to understand a literal expression is shorter than for metaphors, which are equally fast as anomalies**
6. **There is no electrocortical 'shock' for metaphors (cf. anomaly view, Chapter 2)**
7. **Context has no effect on 1 to 6**

Notes:

1. For example:

We are deep in the rain like small insects in high grass (Wij zijn diep in de regen als kleine insekten in hoog gras). Elburg, J.G. (1958). Niemands huis. *Hebben en zijn*. De Bezige Bij, Amsterdam, 12.

The sun plays at my feet like a serious child (De zon speelt aan mijn voeten als een ernstig kind). Rodenko, P. (1975). Februarizon. *Tulpensnijder, orensnijder, verzamelde gedichten*. De Harmonie, Amsterdam, 36.

The day opened like an oyster shell (De dag ging open als een oesterschelp). Snoek, P. (1982). Georgië 1. *Verzamelde gedichten*. Manteau, Antwerpen, Amsterdam, 121.

2. A poetic syllogism has the form indicated below:

A swan is white		$A = B$
Everything white is a light	$(\forall x)$	$B = C$
Therefore, the swan is a light	$\rightarrow$	$A = C$

Notice that a poetic syllogism does not necesserily claim truth value. The example is based on an example drawn from D.L. Black (1990: 212).

3. Although Skinner is not known for this, in *Verbal Behaviour* (1957) he tried to give a full account of ordinary and literary language with the tripartite relation stimulus-response-reinforcement. In *The mand in literature* (p. 49), *Multiple causation in literature* (p. 239), *Formal strengthening in prose and poetry* (p. 246), *Style* (p. 282), *Witty and stylistic effects* (p. 305) and *Composition and its effects* (p. 344), Skinner accounted for poetic instruments such as metaphor, metonymy, synecdoche, oxymoron, rhyme, humor, satire, in the work of such different authors as Shakespeare, Milton, Wilde, Joyce, Huxley, Eliot, and Pound. Mand, tact and multiple causation are his tools for the study of ordinary and literary texts. His statistical approach to alliteration is quite similar to the approaches seen in modern empirical studies into literature:

> The customary practice in literary criticism is to demonstrate such formal properties of poetry and prose by pointing to instances. There is justification for this when we consider the effect upon the reader or listener, of whom the critic is an example. But before inferring any process in the behaviour of the writer, it is necessary to allow for the patterning of his verbal behaviour to be expected from chance. In no case, perhaps, can we say that any one instance of alliteration or other formal similarity is due to a special process, but a general pattern may be demonstrated. Alliteration, for example, may be detected by a statistical analysis of the arrangements of initial consonants in a reasonably large sample. A tendency to alliterate is shown by the extent to which the initial consonants in a given literary work are not distributed at random (Skinner 1957: 247).

4. The range between tautology and anomaly may be seen as a continuum from *total similarity* to *total dissimilarity* between the two terms, which may be explained by the decreasing size of the shared feature sets (cf. Malgady & Johnson 1980):

tautology	literal expression	idiom	metaphor	anomaly
total similarity	>>>>>>>>>>>>>>>>>>>>>	>>>>>>>>	total dissimilarity	

CHAPTER 2: THE ANOMALY MODEL*

2.0 Anomaly theory

The anomaly theory of metaphor is a comparison view with certain adjustments. Analogous to the comparison theory, the core assumption is the comparison of the *A*- and *B-term*, and the prominent role of similarity. However, anomaly theory also stresses the importance of *dissimilarity* between the *A*- and *B-term*. An expression is initially perceived as literal. However, if the expression cannot be interpreted literally, it is perceived as an anomaly. In an attempt to rectify this anomaly, a second, figurative, interpretation is commenced, to test if the expression can be interpreted as a metaphor. When the figurative interpretation also fails, the expression remains an anomaly.

The theoretical advantage of the anomaly view over the comparison view is the explanation why an implicit comparison is transformed into a metaphor. The reason is that the implicit comparison cannot be interpreted literally, and therefore, it must be transformed into the 'is like' form to interpret the expression figuratively. Additionally, anomaly theory considers the *effect* that a metaphor may have upon a reader. As a result of being a restored anomaly, a metaphor evokes a kind of 'shock effect'.

2.1 Anomaly theory in literature

Richards is among the first literary scholars of this century to reflect on the consequences of the surrealist's poetics for the theory of metaphor. On the one hand, he rejected the extreme surrealist view that any two semantically unconnected words still form a good metaphor, on the other hand, he stressed the importance of dissimilarities (disparities) in metaphor understanding and the 'baffling' effect it brings about:

> In general, there are very few metaphors in which disparities between tenor and vehicle are not as much operative as the similarities. Some similarity will commonly be the ostensive ground of the shift, but the peculiar modification of the tenor which the vehicle brings about is even more the work of their unlikenesses than of their likeness. (Richards 1965: 127)

Richards, in the above quotation, provisionally describes a two-stage model of metaphor processing. The search for similarities forms the ostensive ground (*C-term*) of understanding, which is the first stage of the model. However, the number of dissimilarities causes a shift of meaning, so that the

* Notes are on the pages 42-44.

expression is interpreted in a new way, which forms the second stage of the model. Which features, then, are compared in the first stage of understanding a metaphor, and which in the second? It may be assumed that the features of the second stage are responsible for the final understanding of the metaphor:

> Whether, therefore, a word is being used literally or metaphorically is not always, or indeed as a rule, an easy matter to settle. (...) If we cannot distinguish tenor from vehicle then we may provisionally take the word to be literal; if we can distinguish at least two co-operating uses, then we have metaphor. (Richards 1965: 119)

Two uses of a word (or expression) are distinguished by Richards: A literal and a metaphorical one. If tenor (*A-term*) and vehicle (*B-term*) cannot be distinguished, the expression is literal. In other words, if *A-* and *B-term* cannot be identified as different, then they must share many features. The literal interpretation of an expression would then be caused by the almost complete overlap of features between *A-* and *B-term*. As argued above, this is the first stage of the process. Further, the metaphorical interpretation of the expression would follow from 'two co-operating uses'; these are the literal and the metaphoric use. Metaphoric use can be seen as the activation of the figurative features of *A-* and *B-term*, and thus, understanding the metaphor in the second stage of the process would be accomplished by the comparison of the figurative feature sets of *A-* and *B-term*. On the effect this shift from literal to figurative meaning causes, Richards states:

> The mind is a connecting organ, it works only by connecting and it can connect any two things in an indefinitely large number of different ways. (...). As the two things put together are more remote the tension created is, of course, greater. That tension is the spring of the bow, the source of the energy of the shot, but (...) bafflement is an experience of which we soon tire, and rightly. (Richards 1965: 125)

Therefore, Richards favors a milder form of anomaly theory, and opposes the extreme form, represented by André Breton and Eastman, the latter whom he quotes:

> The poet communicates a kind of experience not elsewhere accessible. The poet must arouse a reaction and yet impede it, creating a tension in our nervous system sufficient and rightly calculated to make us completely aware that we are living something - and no matter what. (Eastman, *The Literary Mind*: 205, quoted in Richards 1965: 124)

Eastman represents the extreme of anomaly theory, which shows a certain resemblance with an arousal theory. Nervous tension in the brain would lead to higher cortical activity in response to a metaphor.

Anomaly theory is mixed up easily with comparison theory, because some of their basic characteristics are identical. In the classification of metaphor theories by Mooij (1976: 37), for instance, Henle was considered among the comparison theorists. Indeed, like Richards, Henle (1966) saw the perception of similarity as an important feature of metaphor, and to this extent he fitted in with the comparison view:

> In the metaphor, it is not merely that there are parallel situations - the same elements in the same arrangement, but also that there is a felt similarity between corresponding components. (Henle 1966: 180)

It is also correct that Henle (1966: 173) used Aristotle's theory of metaphor as a starting point. On the other hand, Henle adjusted the comparison view, by making the idea of the 'duality of sense' more precise. Henle stated:

> Each of these words appears in a double role - first in its conventional sense such as it might have in other contexts and second in a sense characteristic of this metaphor. This is what is central in Aristotle's statement.
>
> This duality of sense is characteristic of metaphor and some terminology will make reference to it easier. By the *literal sense* of a word we may mean the sense which a word has in other contexts and apart from such metaphoric uses. By *figurative sense* we may mean that special sense on which the metaphor hinges. (Henle 1966: 174)

This precision in how the double meaning of a metaphor is understood is a main characteristic that distinguishes comparison theory from anomaly theory. In Aristotle's view, no distinction was made between literal and figurative senses of a metaphor; the interpretation of a metaphor is merely a transfer of meanings (features), or the finding of common features, irrespective of literal or figurative meaning.

According to Henle, however, the literal sense of a metaphor is 'the meaning of a term given by a dictionary'. And, 'it is only through the literal sense that one arrives at the figurative' (Henle 1966: 175). This means something else than just finding similarity. Henle's statement implies that two attempts are made to understand a metaphor. First a literal interpretation is undertaken, before the figurative meaning of the metaphor is captured. In the comparison view, no second attempt is made, because the understanding of a metaphor is merely based on finding sufficient overlap to make the expression figurative. Ranking Henle's ideas among the comparison theories - as Mooij did - can only be justified by neglecting the difference between a one-stage and a two-stage model of metaphor understanding.

Rather, in a line with Richards, Henle emphasizes that the most important stimulus for making a second interpretive attempt is the psychological shock effect:

> (...) metaphor may be considered from the point of the listener. Here its outstanding characteristic is the sort of shock which it produces. Ordinarily one takes words in their literal sense and this is impossible in a metaphor. This impossibility in fact is what drives one on to seek a figurative sense. (Henle 1966: 182)

More explicitly than Henle, another theorist of literature - Beardsley - turned away from comparison theory to anomaly theory. He advocated the 'verbal-opposition theory'. In this theory:

> (...) no such importation or comparison occurs at all, but instead a special feat of language, or verbal play, involving two levels of meaning in the modifier itself. When a predicate is metaphorically adjoined to a subject, the predicate loses its ordinary extension because it acquires a new intension - perhaps one that it has in no other context. And this twist of meaning is forced by inherent tensions, or oppositions, within the metaphor itself. (Beardsley 1982: 264)

The verbal-opposition theory - as Beardsley called it - can be seen as a form of anomaly theory. The core assumptions of his theory are couched in the following three points. First, there are two levels of meaning: The ordinary extension and a new intension. Second, there is a change from the one level of meaning to the other, the 'twist of meaning'. Third, the 'twist of meaning' is forced by inherent tensions or oppositions.

Now what is the ordinary extension and the new intension? According to Beardsley (1982: 265), the ordinary extension of a word is 'what is generally true of the *objects*'; 'objects' here meaning all the concrete things in the world. By way of comparison, this is corresponding to what Henle calls 'the meaning of a term given by a dictionary' or 'the literal sense' (Henle 1966: 174). The new intension is formed through the connotations of the word. These 'not common accidental features of the objects denoted' can also include the false features of the object, the 'beliefs' about it (Beardsley 1982: 265). Beardsley did not state whether these 'marginal meanings' (Beardsley 1982: 266) are the same as the figurative meanings that Henle distinguishes. A clue to an answer can be found in the following lines:

> (...) we must look for the metaphoricalness of the metaphor, so to speak, in some sort of conflict that is absent from literal expressions. (Beardsley 1982: 269)

Although tautological, the metaphor would have some 'metaphoricalness', and it stands in some sort of conflict with the literal sense of an expression. Thus, metaphorical is not literal. Moreover, if an expression is the contrary to literal, it is, in a word, figurative.

What did Beardsley (1982: 264) mean by 'inherent tensions', which - as in Henle (1966: 182) - are the driving force behind the second interpreting attempt? Beardsley discussed Henle's view on readers' responses in this respect. He stated about Henle:

> His version of the verbal-opposition theory, however, is described in terms of the reader's response - his 'shock' at the 'clash of meanings'. I prefer to state the theory as a theory not about the effect of metaphor, but about the linguistic structure that causes the effect - about the 'clash of meanings' itself. (Beardsley 1982: 267)

The important thing about Beardley's statements on Henle is that the 'inherent tensions' are the same as Henle's 'shock', caused by the 'clash of meanings'. The twist of meaning Beardsley (1982: 264) mentioned, is forced by tensions or oppositions. According to Beardsley, these oppositions are created by the mismatch of property sets:

> (...) the possibility of the metaphorical performance (...) depends on a felt difference between two sets of properties that (...) are taken to be necessary conditions for applying the term correctly in a particular sense (these are the defining, or designated, properties, (...)); second, those properties that belong to the marginal meaning of the term, (...). I said that when a term is combined with others in such a way that there would be a logical opposition between its central meaning and that of the other terms, there occurs that shift from central to marginal meaning which shows us the word is to be taken in a metaphorical way. It is the only way it can be taken without absurdity. (...) The logical opposition is what gives the modifier its metaphorical twist. (Beardsley 1982: 270)

Apparently, a logical opposition is the main factor that urges the reader to understand an expression as a metaphor. What are these logical oppositions, and how are they formed?

Beardsley distinguished two sets of properties (or features), which are activated by a term (or word). One set consists of the defining, or designated features, in other words, the literal ones. The other set consists of the marginal properties, or, as argued above, the figurative features. As another implicit assumption of anomaly theory, it follows that a metaphor of the form '*A* is *B*' activates four feature sets: A literal and a figurative set for both *A*- and *B-term*.

Although Beardsley (1982: 264-266) strongly argued against a comparison view of metaphor understanding, he used the comparison idea implicitly. If a metaphor is read, Beardsley argued, a logical opposition is formed. This means that the two literal feature sets - which describe the 'real', 'logical', world - probably do not match. If they do not match, it follows that they have been compared before the logical opposition could occur. It also follows that not-shared features (enhancing dissimilarity) can be identified only if a search for shared features preceded.

The logical opposition is formed, when too many literal features are outside the shared set. This 'absurdity' can be reduced, if the marginal, or metaphorical (i.e. figurative) feature sets are activated, and the expression can be understood as a metaphor. Thus, logical oppositions are formed by the distinctive features in the literal feature sets of *A-* and *B-term*. In addition, dissimilarity is the result of comparing the feature sets in a search for common features, which are not found after all.

In anomaly theory, the role of context in understanding a metaphor is limited to an effect 'within the expression'. The shift from one meaning to another - from literal to figurative - could be seen as a context effect of the *B-term*. Only seldom, anomaly theory focuses on the function of surrounding text, and if so, no attempts are made to explain how this text influences the shift. Anomaly theory does not clarify the problem whether the size and contents of the literal or figurative overlap changes because of larger context. Moreover, if the anomaly claim is true that the stage of literal interpretation always precedes the stage of figurative interpretation, context should not have any influence on the serial order of this process.

2.2 **Anomaly theory in psychology**

According to the anomaly theory, metaphors are processed in two stages. First a literal interpretation of the expression is performed, and if it fails, a figurative interpretation takes place. Therefore, the literal, 'normal', procedure would take less time to understand than the figurative.

> (...) all semantically unacceptable sentences may be interpreted as metaphors. This, of course, does not imply that such sentences always, or even regularly, will receive a metaphorical interpretation. A metaphor in this view is, therefore, a semantically unacceptable sentence that has been inferentially interpreted. The interpretation consists of elaborating the metaphor in such a way that an explicit comparison is being made, involving only semantically acceptable sentences and replacing the original unacceptable sentences. (Kintsch 1974: 37)[1]

Unlike the theory of literature, psychology has frequently tried to test the anomaly model (e.g., Harris 1976; Glucksberg, Gildea & Bookin 1982).[2]

These tests usually focused on the time relations predicted by the anomaly view, dependent on whether the literal and figurative stages could be represented by a serial fixed-order model or by a model of parallel stages. Ideas about the contents of these stages were adopted from linguistics, or the theory of literature. A literal interpretation of a metaphor would involve the activation of features describing the actual, 'real' world,[3] whereas figurative interpretation would reflect the activation of the symbolic, imaginitive features of the metaphor (for a linguistic discussion on this matter, see Section 2.1).

Ortony (1979) tried to explain literal features in terms of high-salient or high-frequent predicates, and figurative features in terms of low-salient or low-frequent predicates. Ortony used this frequency criterion for the classification of literal comparisons, metaphors and anomalies. In literals, similarity is thought to be the result of matching high-frequent features; in metaphors, high-frequent *B-term* features would match with low-frequent *A-term* features, and in anomalies, no high-frequent feature of the *B-term* would match an *A-term* feature at all (Ortony 1979: 194-196).

Since psychology tested the presumed two stages in the anomaly theory usually by means of reaction time experiments, these stages were never represented by a feature activation and comparison model. The model presented in Section 2.3 provides for this neglect. Based on the assumptions of the nature of the activated feature sets, it follows from the theory of literature that in the literal interpretation, literal features - initially evoked by *A-* and *B-term* - are compared to form a literal *C-term*, after which figurative, or symbolic, features are activated and compared in the figurative interpretation to form a figurative *C-term*. Thus, while processing the metaphor, four feature sets are activated and compared, namely a literal and figurative feature set for *A-* and *B-term*. As argued in Section 2.1, anomaly theory does not mention whether context has an effect on size or contents of the feature sets. If the claim is valid that the stage of literal interpretation always precedes the stage of figurative interpretation, context should not have any influence on the serial order of this process.

In the theory of literature, the anomaly view on metaphor processing assumes that metaphor evokes some sort of shock effect (see the discussion on Richards, Henle and Beardsley in Section 2.2). In psychology, this view is sometimes translated into terms of arousal-potential. For instance, Anderson (1964) proposed a cognitive model based on the assumptions of Berlyne's arousal theory.

In his discussion of Berlyne's (1960) theory, Anderson related the effect of metaphor with a special cortical activity. The quotations below are drawn from Anderson. The page numbers are Anderson's references to Berlyne:

> Berlyne demonstrates that certain stimulus variables (novelty, "surprisingness", incongruity, ambiguity and uncertainty, and complexity) activate the orienting reflex, or vigilance (pp. 83-86) and

> thereby induce a degree of arousal that forces the individual to reduce that arousal back to a tolerable level. This maintenance of an optimal arousal or steady state (p. 200) is carried out largely by cortical activity (p. 181), although the nature of this interaction between cortex and arousal system has still to be settled.
>
> According to Berlyne's model, all reinforcement is primary and takes the form of arousal-reduction; the organism has neither a need for, say, good nor for achievement, but only for a reduction in the arousal that it has learned to associate with absence of good and of achievement. Occasionally, the organism will seek incentives with arousal value, incentives which will provide "arousal jags" (pp. 198-199), but only provided the arousal-increase is moderate and has a high probability of being followed by arousal-decrease. (Anderson 1964: 59-60)

This motivational theory describes the need for the human cortex to be moderately excited. Anderson puts this idea in use for an assumption originating from surrealist art theory, that metaphor evokes a mild form of 'bafflement' (see Section 2.2). Anderson thus procedes:

> The converse of the arousal jag is boredom, which occurs when there is a rise in arousal with no available means of arousal-reduction. In the case of the experiments on sensory deprivation, the rise in arousal level is explained as a product of the fact that inhibiting impulses from the cortex, which normally dampen arousal, are inactivated as a result of the monotonous stimulation, an event that releases the arousal system from cortical restraint (p. 191).
>
> According to this motivational model, a desirable incentive is an incentive that can provoke an arousal jag: i.e. an incentive that can induce arousal and guarantee its immediate reduction. The metaphor can certainly be categorized in this way. By its linkage of normally disparate states or events, the metaphor is characteristically novel and incongruous; hence it induces arousal. And by the provision of some conceptual resolution in the opposite and appropriate use of this blend of the disparate, it sets off arousal-reduction. (Anderson 1964: 60)

In this view, metaphor is supposed to evoke mild cortical excitement, which should be absent in literal statements and should be paramount during the reading of anomalies.

2.3 A formalized model of the anomaly theory

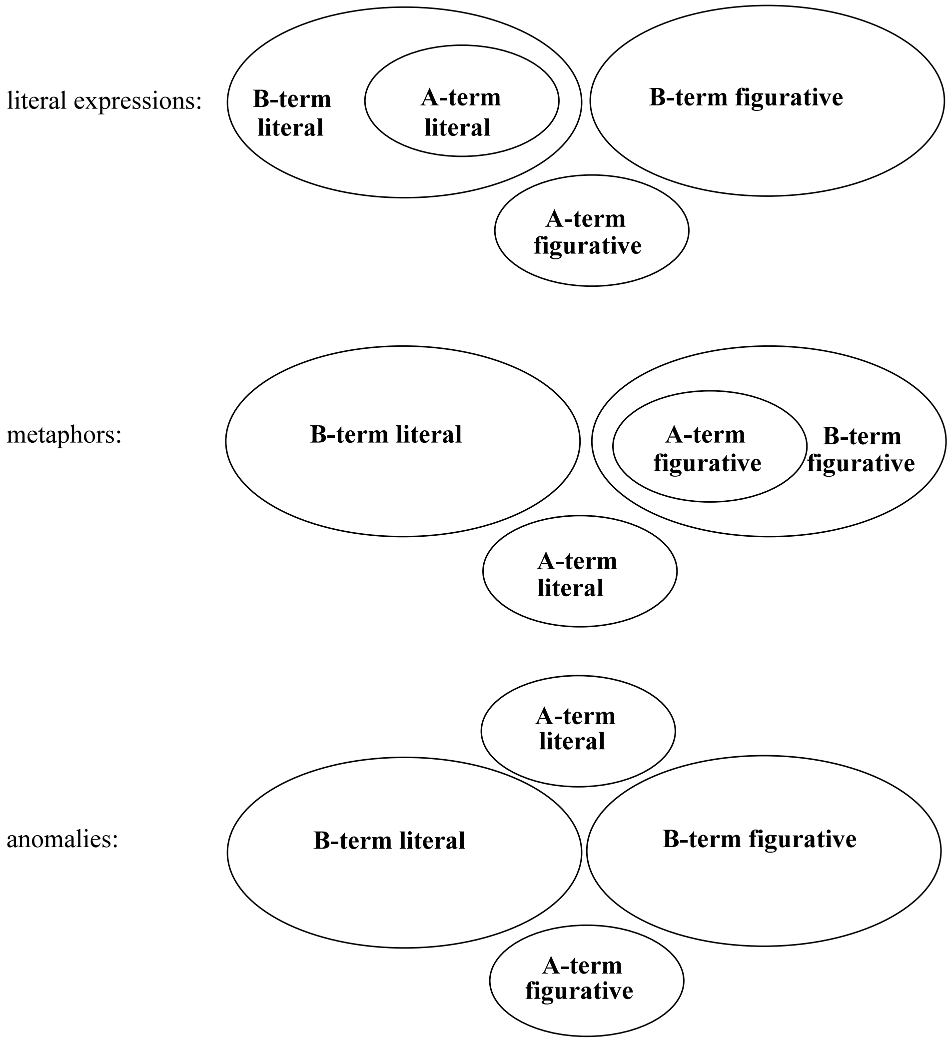
From the above review (Section 2.1 and 2.2) it follows that the core assumptions of anomaly theory are:

1. A metaphor is not a literal statement
2. The surface form of a metaphor consists of two terms
3. A metaphor is understood in two steps
4. In the first step each term activates a literal feature set
5. In metaphor, the number of shared features - the *C-term* - between the literal sets is insufficient
6. In case of a metaphor this causes an absurdity
7. Therefore, two extra, figurative, sets are activated and matched to find the shared set
8. In metaphor, the number of shared features - the *C-term* - between the figurative sets is sufficient
9. Context has no effect on 1 up to 8

Anomaly theory relies heavily on the distinction between literal and figurative features. Literal features are viewed as the concrete, describing properties. For instance, 'fist' may evoke the features 'thumb', 'fingers', 'knuckle', 'ball of the thumb', 'palm' and 'nails'. Figurative features of 'fist' may be 'anger', 'resistance', 'strength', 'aggression', or even 'strike' and 'demonstration'. They are symbols, sometimes drawn from personal fantasy.

Anomaly theory claims that a metaphor will be read first as a literal expression. If the overlap between literal features is insufficient during the literal reading - as can be expected of metaphors - the expression is considered invalid or anomalous. To reduce the relative number of distinctive (literal) features, a figurative interpretation is performed, which involves the matching of two new (figurative) feature sets.[4] If the relative dissimilarity cannot be reduced - because too little a degree of figurative similarity is found - the expression is considered anomalous. Since the extreme form of anomaly theory claims that the stage of literal interpretation precedes the stage of figurative interpretation, context should not have any effect on the serial order of this process.

Figure 2.0: Anomaly model for a four sets feature overlap in literal, metaphoric and anomalous expressions.



It is reasonable to assume that the initial literal interpretation of an expression is based on the literal features of the *A*- and *B-term*. Readers will try to find common literal features first, and only if it is impossible to find them in sufficient numbers, figurative features are activated. In the case of a metaphor, insufficient literal features may be found to decide in favor of a literal expression. When the overlap of literal features is too small for a literal interpretation, a figurative interpretation is performed, in which the distinctive literal features are neglected, and figurative features emerge. It also seems reasonable to assume, that the figurative features constitute the figurative interpretation.

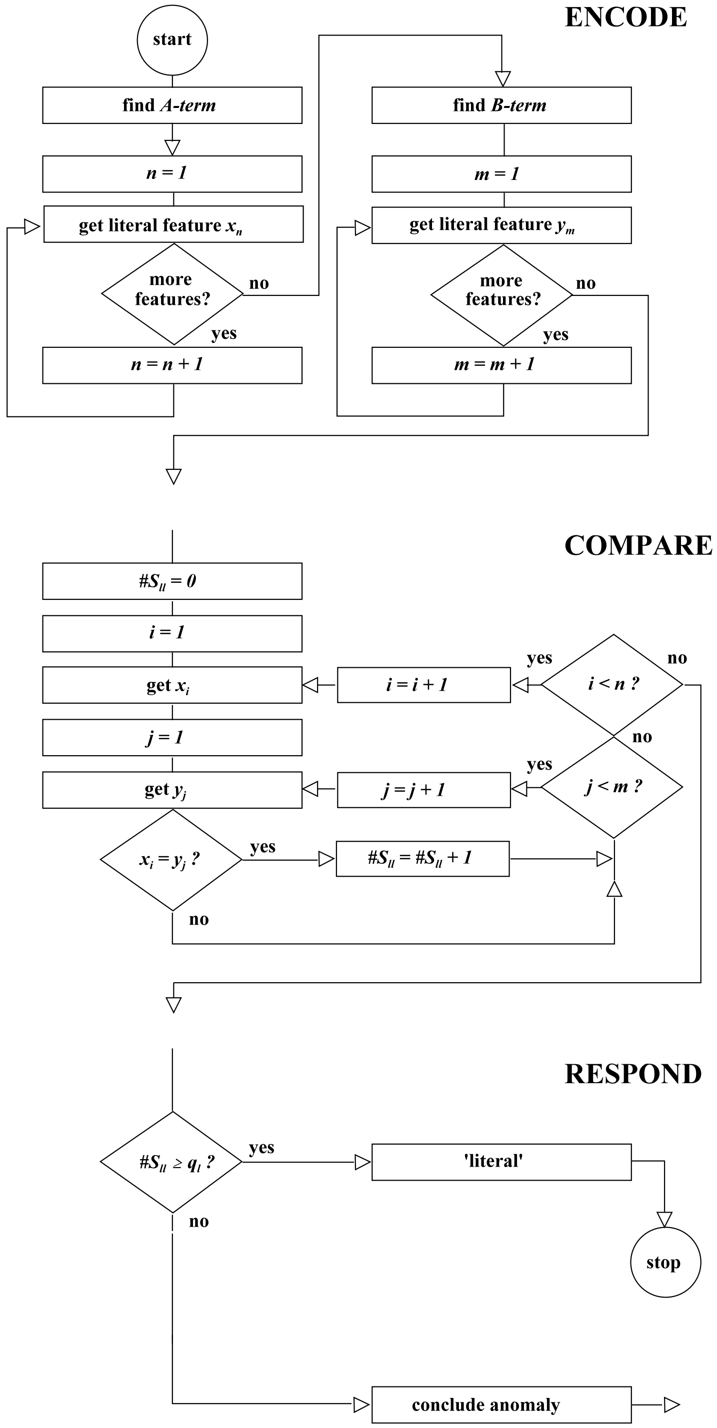
However, the anomaly theory may be interpreted in two ways. First, the figurative feature overlap may be additional to the literal. Due to literal and figurative overlap, the ratio between shared and distinctive features is positively influenced, so that the expression is interpreted as a metaphor. In this case, similarity in an expression is estimated as a measure of overall feature overlap. However, in that case, anomaly theory would merely restate a principle from comparison theory, while loosing a distinctive characteristic.

Anomaly theory would really differ from the comparison view, if literal overlap was not passed on to the later stage. In the second stage, then, the shared and distinctive literal features are neglected, and the feature sets of the *A*- and *B-term* are newly declared with figurative features, the sole overlap size of which would be enough to make the expression metaphoric. Anyhow, the anomaly view implies that every expression activates four feature sets. A literal and a figurative set for the *A-term*, and a literal and a figurative set for the *B-term*.

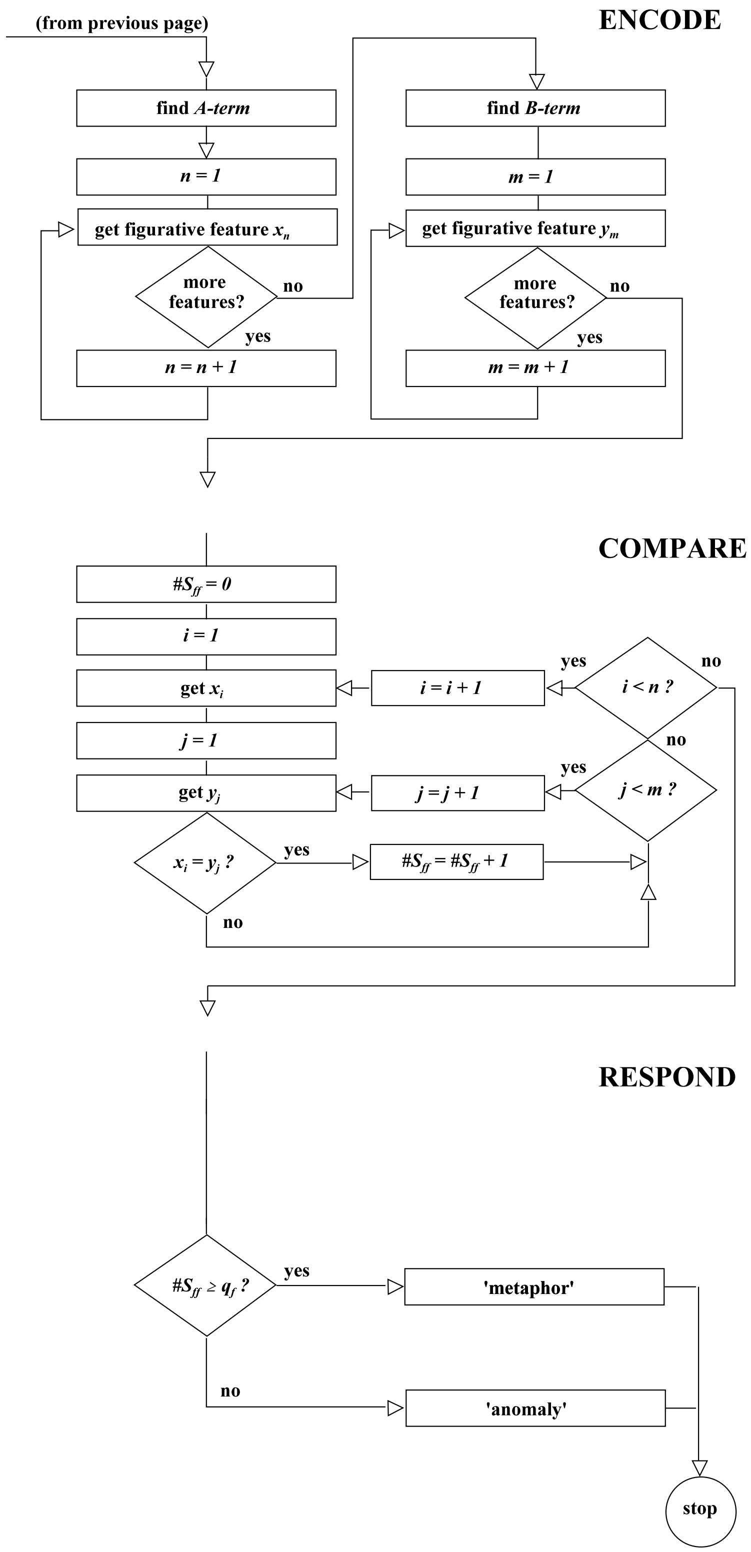
Figure 2.0 shows the Venn diagrams for literal, metaphoric and anomalous expressions, according to the anomaly view. In a literal expression only the literal features are activated and compared, the overlap of which is sufficient to understand the expression. In its extreme form, the complete literal feature set of the *A-term* is found in the literal set of the *B-term* (cf. a tautology). The figurative sets are not compared in literal expressions, and thus, show no matches. In a metaphor, the understanding of the expression takes place only through the figurative features. In its extreme form, the figurative set of the *A-term* is completely matched by the *B-term*. The feature overlap of the literal sets, therefore, should be very small or even absent. In an anomaly, neither literal nor figurative feature sets provide sufficient overlap to understand the expression. In its extreme form there is no feature overlap at all.

Figure 2.1 shows a two-stage serial process anomaly model. The comparison model, as presented in Section 1.3, can fully represent the anomaly assumptions, provided that some adjustments are made. The principal point is that the stages in the anomaly model are both an

Figure 2.1: Two-stage serial process anomaly model for metaphor understanding (fixed order).



(go to next page)



(from previous page)
ENCODE
find A-term
n = 1
get figurative feature x_n
more features?
no
yes
n = n + 1
find B-term
m = 1
get figurative feature y_m
more features?
no
yes
m = m + 1
COMPARE
$\#S_{ff} = 0$
i = 1
get x_i
j = 1
get y_j
$x_i = y_j$?
yes
no
$\#S_{ff} = \#S_{ff} + 1$
i = i + 1
j = j + 1
i < n ?
yes
no
j < m ?
yes
no
RESPOND
$\#S_{ff} \geq q_f$?
yes
no
'metaphor'
'anomaly'
stop

Aristotelian comparison and thus, that each stage in the anomaly model may be represented by a comparison model. In the description of the anomaly model, the discussion will be limited to those aspects which differ from the comparison model. For the arithmetical procedures, see Section 1.3.

In the *encode phase* of the first, literal, stage of the anomaly model (Figure 2.1), feature set X_l of the *A-term* is completely filled with literal features. Feature set Y_l of the *B-term* also consists of literal features only.

The number of shared literal features ($\#S_{ll}$) in the *comparison phase* of the anomaly model is calculated in exactly the same way as in the comparison model, but in this stage for the literal features only.

In the *response phase*, the value $\#S_{ll}$ for literal matches is tested against an inner criterion q_l for literal overlap. Criterion q_l corresponds to the feature overlap shown in the top Venn diagram of Figure 2.0, and must be met to judge an expression as literal. If criterion q_l is satisfied, the process stops and no second stage is entered.

If q_l is not met, in other words, if the number of shared features is not satisfactory, too many distinctive features prevail. The expression becomes an absurdity, and is judged as anomalous. The energetic effect (the 'shock') provoked by this anomalous moment urges the reader to enter the second, figurative, stage of the process.

The figurative stage of the anomaly model is a repetition of the first stage, in which two new feature sets X_f and Y_f are activated. Feature set X_f of the *A-term* is now filled with figurative features, together with the feature set Y_f of the *B-term*. Yet, the actual procedure in the brain can be viewed in different ways. The literal features may be activated first, and only when necessary, figurative features are activated, thereafter. Alternatively, literal and figurative features may be activated at the same time, but the figurative features may become dominant, only when the literal ones do not match.

In the second *comparison phase*, the number of shared figurative features $\#S_{ff}$ receives a value, which is tested against the 'threshold' criterion q_f for figurative feature overlap in the second *response phase*. Criterion q_f is linked to the feature overlap in the middle Venn diagram of Figure 2.0 for metaphors. If q_f is met, the initial anomalous expression is understood as a metaphor. If not, the expression is still an anomaly. The values of the criteria q_l and q_f , again remain empirical questions. To arrive at a decision for 'metaphor', inception of a new stage is demanded, thus increasing the process time beween literals and metaphors. This increment is higher from the anomaly than from the comparison point of view. Comparison theory expects less time delay between literal expressions and metaphors, because merely a simple criterion check separates the decisions. In contradistinction, anomaly theory supposes the onset of a complete new stage between the decisions for 'literal' and 'metaphor', which is probably a more time consuming procedure. As in Figure 1.1 (Chapter 1), decision time differences between metaphors and anomalies are not present, but in view of the results in Chapter 6, the order of processing should be revised anyway.

Computation of the two-stage anomaly model, involves the procedures as described below:

1. encode *A-term*: $X_l = \{x_1.....x_n\}$
2. encode *B-term*: $Y_l = \{y_1.....y_m\}$
3. compute $S_{ll} = X_l \cap Y_l$
   if $\#S_{ll} \geq q_l \rightarrow$ respond 'literal' and stop
   else go to 4
4. encode *A-term*: $X_f = \{x_1.....x_n\}$
   encode *B-term*: $Y_f = \{y_1.....y_m\}$
   compute $S_{ff} = X_f \cap Y_f$
   if $\#S_{ff} \geq q_f \rightarrow$ respond 'metaphor'
   else respond 'anomaly'
5. stop

q_l = number of literal elements needed in $X_l \cap Y_l$ to respond 'literal'
q_f = number of figurative elements needed in $X_f \cap Y_f$ to respond 'metaphor'

As suggested in Section 1.3, the model could be refined by using weighted features. This means that some literal or figurative features *u* of a term may be more important than others (e.g., more salient, more prototypical, more frequently mentioned). The weighing of the features *u* could thus lead to a preferred order of comparison within the literal or figurative feature sets. Analogous to Section 1.3 (Chapter 1), the sizes of the weighted literal feature sets X_l and Y_l are described by:

$W_{Xl} = \sum_{u\, \varepsilon\, Xl} w_{Xl}(u)$, and $W_{Yl} = \sum_{u\, \varepsilon\, Yl} w_{Yl}(u)$, respectively.

The sizes of the weighted figurative feature sets X_f and Y_f are described by:

$W_{Xf} = \sum_{u\, \varepsilon\, Xf} w_{Xf}(u)$, and $W_{Yf} = \sum_{u\, \varepsilon\, Yf} w_{Yf}(u)$, respectively.

Again in correspondence with Section 1.3, the shared literal set size $\#S_{ll}$ and shared figurative set size $\#S_{ff}$ can be rewritten as the weighted set size (W) of all shared features u, which is the sum of all weights w of all features u in S_{ll} and S_{ff} , respectively:

$$W_{Sll} = \sum_{u \,\varepsilon\, Sll} w_{Sll}(u) \, , \qquad \text{and} \qquad W_{Sff} = \sum_{u \,\varepsilon\, Sff} w_{Sff}(u) \, .$$

The model in Figure 2.1 suggests that when criterion q_l is not satisfied, a second, figurative, stage is started, before another stop criterion is found. This need not be the case. Subjects could decide that if the expression is not literal, it is thus an anomaly, and cancel all further processing. If so, another decision diamond ought to follow (or be in parallel with) q_l , predicating that after the conclusion for anomaly, the process halts, unless a minimum criterion of overlapping literal features urges the subject to activate figurative ones.

From the model in Figure 2.1, predictions about metaphor processing can be derived, in agreement with the anomaly assumptions. These predictions are competitive with the predictions in 1.3 (comparison model) and 3.3 (interaction model).

The predictions in accordance with the anomaly view are:

1. **A metaphor activates four feature sets**
2. **These four feature sets are always fixed**
3. **A literal expression has more shared literal features than a metaphor and an anomaly; a metaphor has more shared figurative features than a literal and an anomalous expression**
4. **The set of literal and figurative features is fixed, and thus, can be predicted from the sets of the isolated words**
5. **The response time for understanding a literal expression is less than for metaphors and for metaphors equal to that for anomalies**
6. **Metaphors and anomalies show electrocortical effects for semantic mismatches**
7. **Context has no effect on 1 to 6**

Notes:

1. Kintsch's (1974) model on sentence comprehension can be seen as an anomaly theory, as well as an interaction theory. Metaphor is a semantically unacceptable expression (an anomaly), which is resolved by 'inferences', a notion of the interaction theory (cf. 'analogical reasoning'):

> The model for semantic memory, (...), combines the virtues of network models (that is, their generative capacities) with those of set-comparison models (that is, their flexibility). The measure of semantic distance derived from the model is a function

> of both the amount of inferential processing that is performed in comparing two concepts and their element overlap (Kintsch 1974: 211).

Metaphor understanding involves simple feature matching as well as inferencing. This brings Kintsch close to Gentner's (1989) position, who claimed that metaphors are constituted by shared attributes and shared analogical relations (see section 3.2).

To treat Kintch's theory as an anomaly model is rather arbitrary, of course. However, this decision is based on the idea that Kintsch's ultimate ground to calculate distance is still feature overlap, and that inferences are weighted less over time:

> It is assumed that in calculating an over-all semantic distance measure subjects combine overlap and inferential information so as to weigh inferred information less and less (Kintsch 1974: 212).

Kintsch then developed a two-stage model for sentence processing:

> Stage 1 processing consists first of (...) translating from sentence into proposition. Next a check is made on whether the test proposition is already stored in semantic memory, in which a fast "true" response occurs. (...) If the proposition is not found in memory, the semantic distance between its arguments is computed. If this distance is zero or close to zero, a fast "false" response occurs, as in the case of the nonsense sentences (...). If, however, the semantic distance is above a critical value, further processing is necessary. (...) this may be very time consuming (...). In principle Stage-2 processing consists of determining whether or not the test proposition is semantically acceptable (...). The model is incomplete in that it does not fully specify how this is done; it merely claims that the production rules that are part of a person's semantic memory are used to generate inferences from the material directly stored in memory (Kintsch 1974: 212-213).

Despite the problem that the model does not account for the determination of semantical acceptability in Stage-2, it suffers from other problems as well. In Stage-1, a full set of already stored propositions is assumed. First of all, this means that the brain conserves an exhaustive set of fixed word combinations (the propositions), which must be checked completely, before the 'true' response occurs. Thus, the older one gets, the longer it takes to check the set completely. Besides, the brain is known to remember only the relevant parts of a sentence and to forget surface forms, in this case, the fixed propositions. Kintsch tried to account for this problem by assuming a noun-hierarchy deletion operator, which deletes redundant propositions (Kintsch 1974: 26). However, this deletion operator was never investigated experimentally.

The model Kintsch advocated can hardly explain the quick understanding of newly heard sentences, because the sentence has to be translated into a proposition, which must be checked against a set of already learned propositions, after which a distance is calculated, which, in turn, must meet a certain critical value. Although parsimony is not an argument to Kintsch (1974: 202), a much simpler model presented in Section 2.3 can account for Kintch's assumptions as well.

2. The anomaly model is sometimes referred to as the Standard Pragmatic Model (cf. Gerrig 1989_a).

3. In this connection, the Clark & Lucy (1975) study into indirect requests speaks of literal meaning as lexeme meaning.

4. Steen (personal communication) stressed the point that anomaly theory does not envision two figurative sets in the second stage, but only one for the *B-term*, so that matches occur between the literal *A-term* set and the figurative *B-term* set (cf. Beardsley 1982: 264-

265).

Although strictly speaking, this is a valid comment, the models try to make the sharpest distinctions possible among the theories. The assumption of three feature sets would intersect with three sets for the interaction theory (next Chapter). Moreover, the results in Chapter 5 will show that the model should be revised in favor of, indeed, matches between literal and figurative features. However, these are not only matches from the figurative *B-* to the literal *A-term* set, but also from the figurative *A-term* to the literal *B-term* set. In other words, the four feature sets assumption is essential to explain the results.

CHAPTER 3: THE INTERACTION MODEL*

3.0 **Interaction theory**

Interaction theory is the most recent of the dominant views in the theory of literature. It has the most advanced, but also the most vague assumptions of the three competitive theories. Similarity and dissimilarity between *A*- and *B-term* are equally important for the processing of a metaphor. In addition, interaction theory claims that the search for similarity is not the most important mechanism. Above all, reasoning by analogy is the way in which meaning is created for a metaphor. The reader tries to invent a relationship between the *B-term* and one of its activated features. This relation is then transferred to the *A-term*. In the *A-term* set, a feature is sought that fits this transferred relation, supposedly resulting in a reasonable solution to the metaphor. Some examples in the Sections 3.2 and 3.3 will illustrate this point.

3.1 **Interaction theory in literature**

In *Poetics*, Aristotle discriminated between two ways in which a metaphor may be understood. By searching for similarities (Section 1.1) or 'by analogy':

> Metaphor 'by analogy' is a case where the relation of *b* to *a* is the same as that of *d* to *c*: the poet will use *d* instead of *b*, or the reverse. Sometimes they add to the metaphor something to which it is related (...).
>
> For some analogies there is a missing term, but the metaphor will still be used. (Halliwell 1987: ch.21, 55-56)

Interaction theory is greatly indebted to this idea. The 'founding father' of the interaction view, M. Black, spoke of a projection mechanism that maps a set of 'associated implications' from the *B-term* upon the *A-term*. This is 'a set of what Aristotle called *endoxa* - current opinions shared by members of a certain speech-community' (M. Black 1979: 29). M. Black stressed that this set may be 'novel and nonplatitudinous' (ibid.). The mechanism he offered, looks like this:

> A metaphorical statement has two distinct subjects, to be identified as the "primary" subject and the "secondary" one. (...). The secondary subject is to be regarded as a system rather than an individual thing. (...). The metaphorical utterance works by "projecting upon" the

* The Note is on page 61.

> primary subject a set of "associated implications," comprised in the implicative complex, that are predicable of the secondary subject. (...) The maker of a metaphorical statement selects, emphasizes, suppresses, and organizes features of the primary subject by applying to it statements isomorphic with the members of the secondary subject's implicative complex. (...) In the context of a particular metaphorical statement, the two subjects "interact" in the following ways: (a) the presence of the primary subject incites the hearer to select some of the secondary subject's properties; and (b) invites him to construct a parallel implication-complex that can fit the primary subject; and (c) reciprocally induces parallel changes in the secondary subject. (M. Black 1979: 28-29)

To summarize the theoretical considerations outlined above, M. Black maintained that the *B-term* (the 'secondary subject') activates a set of 'associated implications'. These associated implications can be viewed as features connected to the *B-term* by some sort of relation. An example given by Tourangeau & Sternberg (1982) says that 'king' may activate the feature 'empire' and the corresponding relation between king and empire may be 'power'. This relation, 'power', is then linked to the feature set of the *A-term* (the 'primary subject'), which may be 'teacher'. In the *A-term* set of 'teacher', a feature should be found that fits the relation 'power'. This could be the feature 'class' or 'children'. The metaphor 'teachers are kings' is solved, then, by the analogy of teachers having power over their classes, like kings have power over their empires (teachers : classes :: kings : empires).

M. Black went even as far as to state that the relation is 'mapped' upon, rather than added to, the *A-term* set. Here the premise is that the *A-term* has activated the relation, before it can be mapped by the *B-term*:

> I have said that there is a similarity, analogy or, more generally, an identity of structure between the secondary implication-complex of a metaphor and the set of assertions - the primary implication-complex - that it maps. In "poverty is a crime," "crime" and "poverty" are nodes of isomorphic networks, in which assertions about crime are correlated one-to-one with corresponding statements about poverty. (M. Black 1979: 31)

If this mapping is one-to-one, there will be fewer possibilities to find a solution to the metaphor. Not only a fitting feature should be found in the *A-term* set, a similar relation should also be present beforehand, rather than transferred from the *B-term*. This assumption impoverishes the freedom of interpretation for the reader. However, if the assumption is valid, the interpretation is almost automatically cued in the right direction. M. Black's regulation of metaphor understanding is less rigid, when it comes to the number of relations that can be mapped by the *B-term*:

> Let us now idealize the connection between the two implication-complexes (G and M) in the following way: G consists of certain statements, say *Pa*, *Qb*,..., and *aRb*, *cSd*,..., while M comprises corresponding statements P'*a*', *Q*'*b*',..., and *a*'*R*'*b*', *c*'*S*'*d*',..., (where P is uniquely correlated with *P*', *a* with *a*', *R* with *R*', and so on). Then the two systems have, as mathematicians say, the same "structure"; they are isomorphic (...). One important deviation from the mathematical conception is that G is linked with M by a "mixed lot" of projective relations, (...), and not (as typically in mathematical contexts) by a single projective relation. (...) G is precisely what I have called in the past an "analog-model". (M. Black 1979: 30-31)

What may be learned from this quotation is merely that the *B-term* (G) is not linked to the *A-term* (M) by just one relation, but by a 'mixed lot' of many (one-to-one) relations. This improves of course, the freedom of interpretation for the reader. The problem M. Black signaled that *A-* and *B-term* in metaphors are connected by more than a 'single projective relation' can be solved by introducing a procedure in the metaphor model that allows any member of a relation set to be compared with any member of the other set. The *B-term*, then, generates two types of feature sets: A feature set of 'normal' *B-term* properties, which may be called Y, and a feature set of relations, which may be called Y_r . In parallel, let the feature set of the *A-term* be X and the feature set of *A-term* relations X_r . The contents of the feature sets X and Y are the same as in the feature sets of the comparison model.

M. Black's view on the connection between the two sets of relations may then look as follows: $Y_r = \{y_{r1} , ... , y_{rm}\} = \{x_{r1} , ... , x_{rn}\} = X_r$, allowing for at least one relation Y_{rj} to connect with X_{ri} . M. Black stated, however, that this connection is insufficient, if the linkage cannot be made by a 'mixed lot', i.e. if $X_r \cap Y_r \ngtr 1$. The model presented in Section 1.3 provides a procedure, which solves this problem in a simple way, so that any X_{ri} and any Y_{rj} can be compared in any direction: $X_{ri} = Y_{rj}$, if $r_i \leq n$ and if $r_j \leq m$. In other words, relation Y_{rj} can be compared with any relation X_{ri} as long as relation set Y_r is not empty. Similarly, any relation X_{ri} can be compared with any relation Y_{rj} , if feature set X_r is not empty. This procedure is demonstrated in Figure 3.1, Section 3.3.

M. Black suggested that the relations in X_r and Y_r may be identical. However, as argued above, this limits the possibilities for interpreting the metaphor. The interaction view is made more powerful, if a relation is transferred from the *B-term* to the *A-term*, so that one or more features in X can be found that fit the transferred relation. The formal representation of the interaction view can then be simplified by searching a feature x_i so that it fits Y_{rj} . Thus, X_r can be rewritten as X, which leaves the interaction view with three feature sets: One normal feature set for the *A-term* (comparable to the kind activated in the comparison model), and two feature sets for the *B-term* (one normal set of the comparison model type and one relation set). A literal statement will be understood by activating the feature sets as described in the comparison model, since the relation inference is supposed to be special to metaphors.

Unlike comparison theory and anomaly theory, interaction theory seriously considers the role of context. In the comparison view, the similarity of two terms is pre-existing, and merely has to be found. Context is not believed to affect its magnitude. Anomaly theory also does not mention a context effect, other than within the expression (the *B-term* influencing the *A-term*). Interaction theory, on the other hand, really insists on the role of context: "At the level of interpretation, text-understanding gives the key for metaphor-understanding" (Ricœur 1974: 105). Ricœur did not think of text and metaphor processing as a simple succession of meanings, which lead to the proper interpretation. However, he saw it as a construction, a holistic, and therefore creative, process:

> Now a work has to be constructed because a text--especially if it is a literary work--is more than a linear succession of sentences. It is a cumulative, holistic process. (Ricœur 1974: 104)

The reader lacks the knowledge about the writer's intention, so that he has to 'guess'. The way the reader 'guesses' the meaning of a metaphor is even exemplary for the way a literary text is understood:

> (...) As concerns the place of *guessing* in the construction it follows from what we said about the absence of the author's intention as a guideline and the character of a work as a system of whole and parts. We may summarize in this way the corresponding features which are the grounds for the analogy between the explication of a metaphoric statement and a literary work as a whole. (Ricœur 1974: 104)
>
> The explication of metaphor as a local event in the text contributes to the interpretation itself of the work as a whole. We could even say that, if the interpretation of local metaphors is enlightened by the interpretation of the text as a whole and by the disentanglement of the kind of world it projects, the interpretation of the poem as a whole is controlled, reciprocally, by the explication of metaphor as a local phenomenon. (Ricœur 1974: 109-110)

Ricœur assumes that text cues (or 'clues') influence the guess of the metaphorical meaning. In terms of the interaction theory; what has to be guessed is the sort of relation that is activated and transferred (or mapped).

> In both cases the construction relies on the "clues" contained in the text itself: a clue is a kind of index for a specific construction, both a set of permissions and a set of prohibitions; it excludes some unfitting

> constructions and allows some others that make more sense of the same words. (Ricœur 1974: 104)

3.2 Interaction theory in psychology

Interaction theory criticizes the comparison view for its simple ideas about the search for similarity. For instance, Tourangeau & Sternberg (1982) enumerated mechanisms they considered more important, among others 'abstraction' and 'natural association'. Unfortunately, they did not work their handful of alternatives into the model they proposed.

According to Tourangeau & Sternberg, similarity does not play a vital role in metaphor comprehension, and its pre-existence between *A-* and *B-term* should be denied. Instead, the *B-term* forms a frame in which the *A-term* can be newly interpreted. Since the *B-term* features do not immediately have to fit the *A-term* features, a 'reorganization' may be required; a reinterpretation of one part of the metaphor in terms of the other.

These ideas were extracted from M. Black (1979) (see Section 3.1), who posited that the interpretation of a metaphor is not so much the comparison of *A-* and *B-term* on existing similarity, but the renewed construction of *A-* and *B-term* to *create* similarity. In this view, understanding a metaphor strongly depends on the competence of the reader.

To support their claims, Tourangeau & Sternberg cited a study by Ortony (1979), who stated that the feature lists summed up by his subjects showed 25% overlap of salient features (salience indicated by the subjects) for literal expressions, and 1% for similes. Tourangeau & Sternberg explained the low percentage of salient shared features for similes as a result of 'reshaping' the shared set into a new concept. Owing to the fact that this new concept evoked new features, none of the original *A-* and *B-term* features could be recovered in the shared set.

However, it can be objected that the results in Ortony (1979) do not demonstrate that the search for similarity or the matching of features did *not* take place, or at least, was not important for creating meaning. It could be argued that one shared feature is enough to form an interpretation, without the necessity to be a highly salient feature. This means that matching features sets may still be crucial for forming meaning, although the responsible features may not be recovered, and the full interpretation may not be explained by it.

To underscore their point, Tourangeau & Sternberg cited the study of Verbrugge & McCarrell (1977), in which subjects performed a recall task with the *A-term*, *B-term* or *C-term* as cue (*C-term* seen as the solution to the metaphor). The results showed that the shared features served as a prime for memorizing the metaphors equally well as the *A-* and *B-term*. Unfortunately, this study is also hardly decisive. Verbrugge & McCarrell's finding does not support the premise that the *C-term* was *not* formed by shared features of *A-*

and *B-term*. The authors themselves indicated that their data are open to various explanations.

Tourangeau & Sternberg's claim was that reinterpreting a feature from the *A-term* in terms of the *B-term* changes the whole domain of *A-term* features. According to this 'domain interaction' theory, every activated feature evokes a feature set that is included in the comparison. Such a feature maintains a specific relation with the feature set, for instance, when a feature is prototypical for a category. If an expression is identified as a metaphor, the reader searches the *B-term* set for a feature that holds such a special relation with the *B-term*. If this relation is found or created, it is relocated to the *A-term*, where it is used as a search light to find an appropriate completion in the feature set of the *A-term*, thus creating the *C-term*.

However, it is a misconception that such a relation of 'feature-to-feature set' would separate interaction theory from a comparison view, as Tourangeau & Sternberg declared. Such a relation can be seen as a feature in its own right, belonging to a feature set of relations. This relation set - for instance of the *B-term* - may be compared with the relation set of the *A-term* and the matches may establish the *C-term* (the solution to the metaphor). If relations do not match, the relation set of the *A-term* may be changed until they do. 'Teachers are kings', then - the example is by Tourangeau & Sternberg - would transport the relation of 'power' between 'kings' and 'empires' to 'teachers' and 'classes'. In other words, interaction theory also may envision a role for feature mapping - in the form of relations - and for searching the shared (relation) set. This shared set does not need to be fixed and may be created by adjusting the feature sets.

Ultimately, Tourangeau & Sternberg did not deny that *A-* and *B-term can* share features and *can* show similarity, but they did not find it important. Nevertheless, the alternative mechanisms they proposed (abstraction, natural association) were weakly explicated, and did not return in their processing model.

This model, then, states that first, the encoding of an expression takes place. The terms are identified as *A-* and *B-term* and activate feature sets. Next, the features are mapped, in particular, the relation of a *B-term* with its domain is delivered to the *A-term*. This *B-term* relation is then applied to the *A-term* set, which is checked for a well fitting completion (making up the same relation). Subsequently, *A-* and *B-term* are compared with respect to this completion, to verify whether the relations really correspond, after which the response selection is justified. If the answer is 'not metaphoric', reinterpretation takes place (in which stage the process restarts, is not mentioned). If the answer is 'yes', the response is 'metaphor'.

Ortony's (1979) study was used by Tourangeau & Sternberg to sustain the relational thrust of metaphor understanding, although Ortony himself objected to such an interaction point of view:

> (...) relations are no more nor less important to the nature of metaphors than are objects. Both are important in that they constitute the kind of things that tie language to reality, but neither of them are powerful tools for explaining specific linguistic phenomena. (Ortony 1979: 187)

Ortony pointed out that, rather than relations, literal falsehood transforms an expression into a metaphor, thus bringing metaphor understanding back to the tenets of anomaly theory:

> Clearly, part of what is involved in understanding this is the solution of the analogy of the form "X is to ? as Y is to Z." What makes it a metaphor is not the fact that common relations are involved, but rather; the fact that, literally interpreted, the assertion is false. It is false because the relations that are allegedly similar, are not in fact similar at all. (Ortony 1979: 188)

Yet, Ortony was not so radical as it seems from the above quotation. He left room for relations as well as for simple feature sharing, taking a moderate interaction view eventually:

> (...) some analogies are literal analogies and some are not. In either case, they involve alleged similarities - similarities between relations between objects, rather than between objects themselves. (Ortony 1979: 188)

In line with Ortony, Gentner (1988, 1989) advocated an intermediate position in the interaction view. She said that metaphors stretch from mere resemblances (simple similarity in the Aristotelian way) to highly relational commonalities. According to Gentner, many relational predicates and many object attributes are mapped for *literal similarity*. In anomalies, few attributes and few relations are mapped (Gentner 1989: 206). For metaphors, the range varies from mapped attributes to mapped relations, thus forming an intermediate class of expressions between analogy, anomaly and mere appearance (Gentner 1989: 207).

The model Gentner (1989) proposed, is in line with the one presented by Tourangeau & Sternberg, which is a form of structure mapping, or, for that matter, analogical reasoning. The process starts by accessing the 'system' (of relations) of the *B-term*. Then, the mapping between *B-term* and *A-term* is performed, the match evaluated, and inferences are stored in the *A-term*. Only seldom, are commonalities extracted (Gentner 1989: 200).

Finally, an important insight of interaction theory is that feature sets do not need to be fixed and similarity does not need to be present already. Interaction theory also stresses the importance of the linguistic context for the possibilities and limitations of the features to connect with other feature sets.

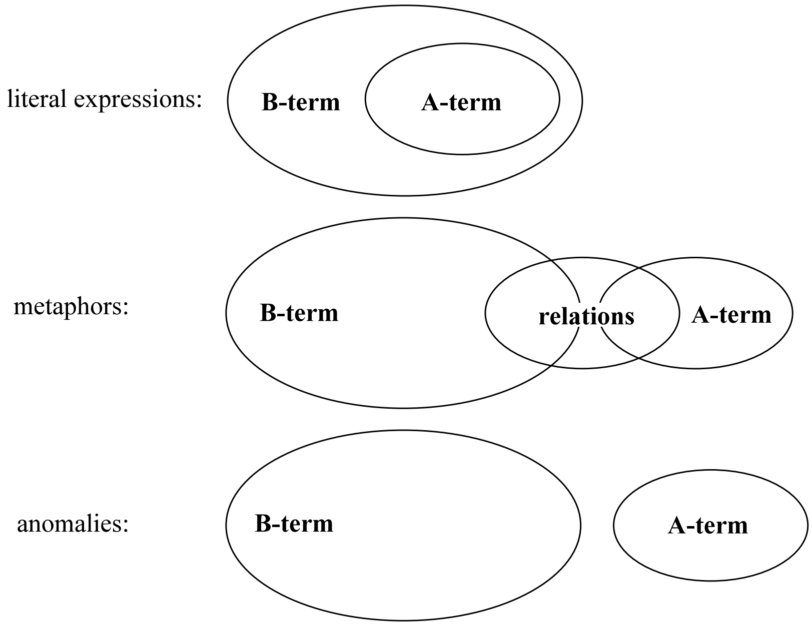
3.3 A formalized model of the interaction theory

Giving the chief points of Section 3.1 and 3.2, the interaction theory assumes that:

1. A metaphor is not a literal statement
2. The surface form of a metaphor consists of two terms
3. A metaphor is understood in two steps
4. In the first step each term activates a normal feature set
5. In metaphor the overlap - the *C-term* - between the feature sets is insufficient
6. An extra relational set is created from the *B-term* and matched for shared relations with the *A-term*, or a fitting *A-term* feature is found for the *B-term* relation, forming the *C-term*
7. In metaphor, the number of shared relations, or the number of relation-fitting *A-term* features is sufficient
8. Context cues influence step 4 up to 7

Where literal expressions are concerned, interaction theory does not deviate from the comparison view. Whether features are literal or figurative is not considered important, and relations do not come into play, unless the feature overlap is insufficient. In that case, a third feature set is activated (or created) for the *B-term*, and mapped onto the *A-term* area. The relation set, therefore, accomplishes a bridge function between the *A-* and *B-term* feature sets. If the relations are appropriate, a metaphor is formed. If not, the expression is perceived as an anomaly. Figure 3.0 shows the Venn diagrams for the activated feature sets, modeled in the interaction way. For literal expressions, the normal feature overlap between first and second term is sufficient. The feature-poor first term is probably fully covered by the feature-rich second term. In metaphor, the overlap is created through relations. Thus, metaphors show one extra set of features. Anomalies share neither normal features, nor relations.

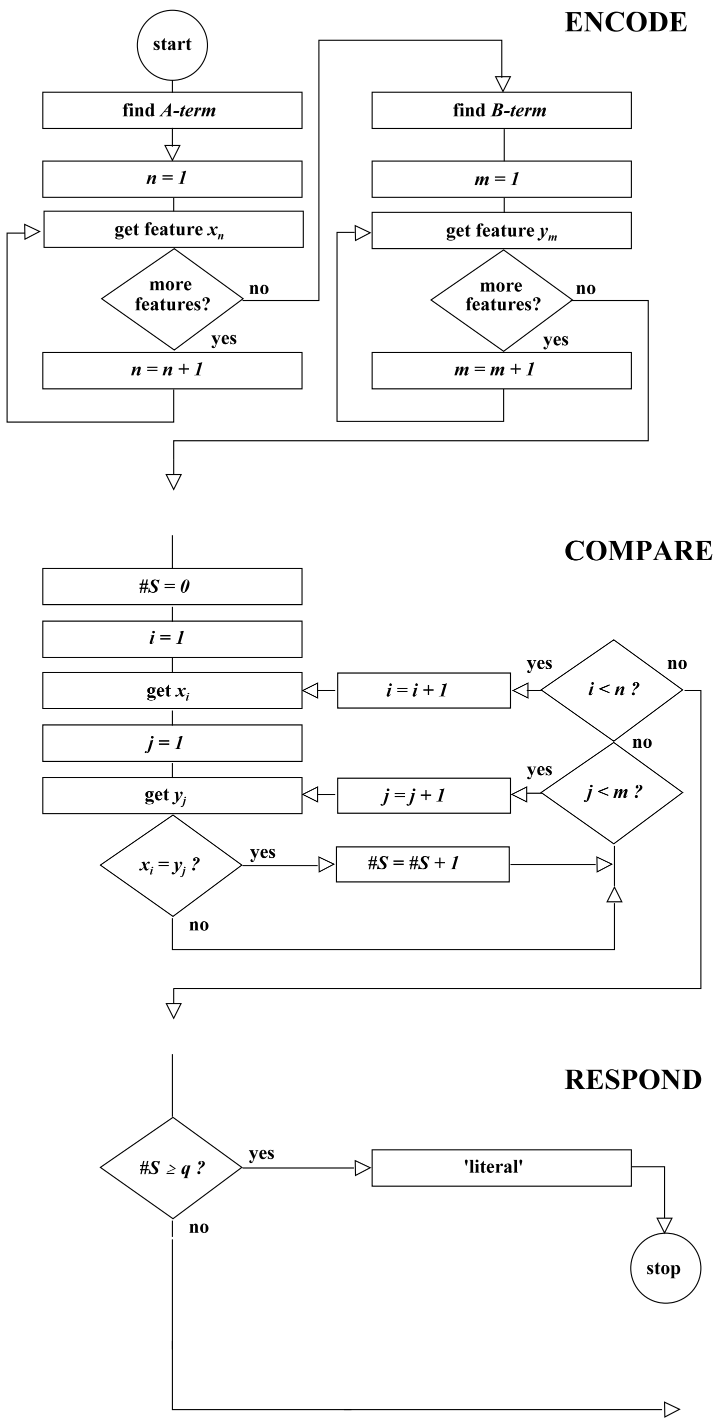
Figure 3.0: Interaction model for a three-sets feature overlap in literal, metaphoric and anomalous expressions.



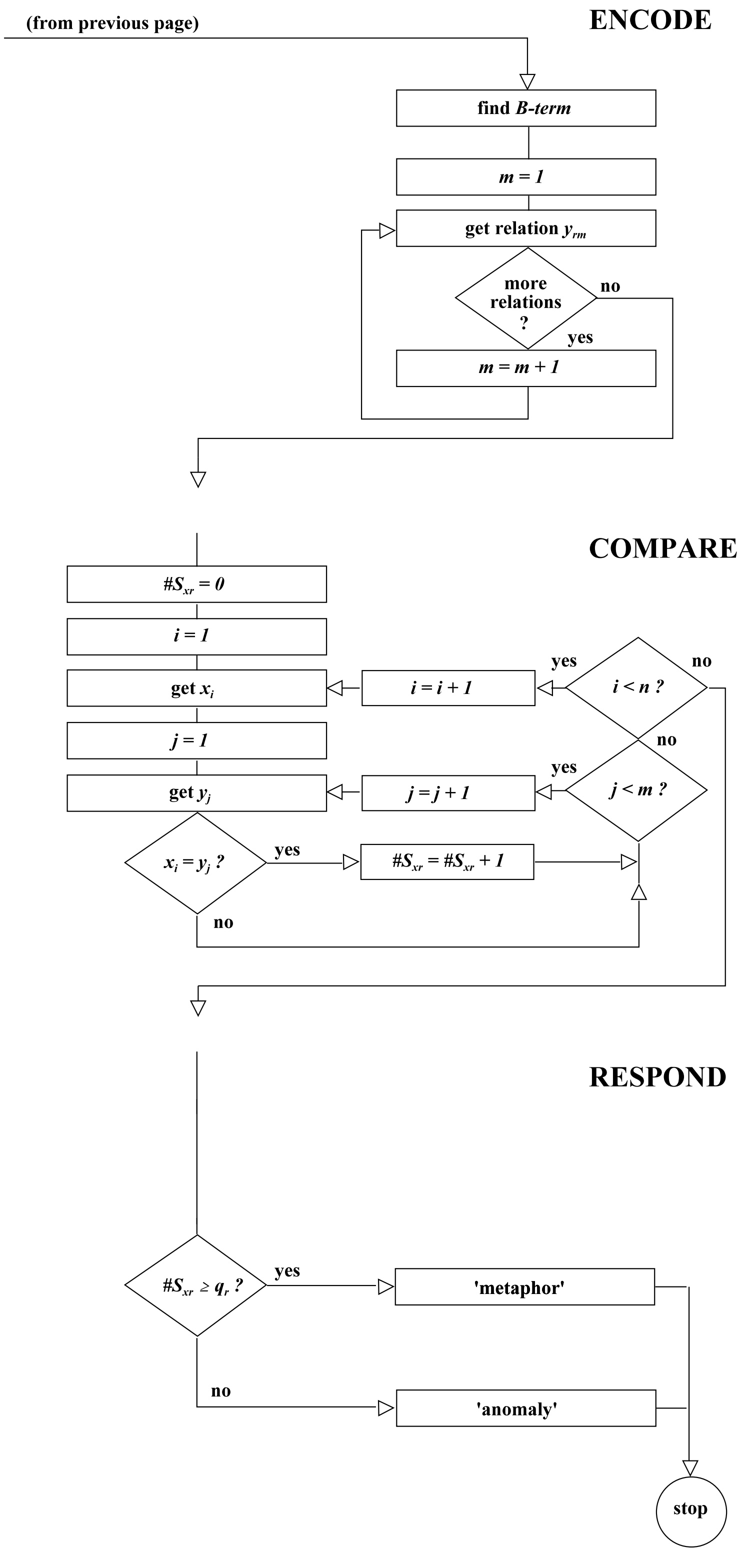
In Figure 3.1 a flow-diagram is presented for a two-stage serial order interaction model. Analogous to the anomaly model in Section 2.3, the stages of the interaction model can be represented by the comparison model, as shown in Section 1.3, Figure 1.1. The two main differences between the anomaly and interaction model are the mixture of features in the first stage of the latter, in contrast with the literal determination of the former. Secondly, the relational shift in the second stage of the interaction model contrasts with the figurative shift of the anomaly model. The presence of a cortical 'shock effect', urging the reader to enter the second stage, is not raised as an issue by the interaction theory.[1]

In the first stage of the interaction model, *A-* and *B-term* evoke features as in the comparison model. Subsequently, they are matched for feature overlap. Below, the description of the first stage in Figure 3.1 is picked up when the criterion *q* is not satisfied (see the comparison model in Section 1.3) and the number of shared features (#*S*) is

Figure 3.1: Two-stage serial process interaction model for metaphor understanding (fixed order).



(go to next page)


(from previous page)
ENCODE
find *B-term*
$m = 1$
get relation y_{rm}
more relations ?
no
yes
$m = m + 1$
COMPARE
$\#S_{xr} = 0$
$i = 1$
get x_i
$i = i + 1$
yes
no
$i < n$?
$j = 1$
no
yes
get y_j
$j = j + 1$
$j < m$?
$x_i = y_j$?
yes
$\#S_{xr} = \#S_{xr} + 1$
no
RESPOND
$\#S_{xr} \geq q_r$?
yes
'metaphor'
no
'anomaly'
stop

insufficient to make a decision for 'literal'. The philosophy of the interaction view on the special role of the *B-term* may now be unscrambled in a model of arithmetic procedures.

In the *encode phase* of the second stage in the interaction model, the *A-term* set remains untouched. A new feature set $Y_r = \{y_{r1}, \ldots, y_{rm}\}$ for the *B-term* is declared, which contains relations among the features in Y.

The procedure in the *comparison phase* can be modeled in two ways, depending on the kind of interaction view taken. In M. Black's view, both X and Y should be supplied with X_r and Y_r in the second stage. In this four-sets model, only a relation match would be appropriate to augment $\#S$.

Figure 3.1, however, models the three-sets view of Tourangeau & Sternberg, who claim that only Y_r is created. A relation y_{rj} is transported to the *A-term*, and a fitting feature in X should complete the relation. How this link is established remains unclear. Tourangeau & Sternberg prefer to speak of 'creation' of meaning, and dismiss the search for shared features as an adequate alternative.

Nevertheless, subjects do not link any relation with any feature in a random guess. There is a decision criterion. This may yet be the size of the overlap between relations and features, causing a higher activation level between these two, than between any other features. Also if the link is understood as a substitution of $(y_{rj} + y_j)$ by $(y_{rj} + x_i)$, the connection may be based on the overlap between the meanings of y_j with x_i , making them exchangeable. In other words, the value of $\#S$ could be calculated as the (partial) identity of a relation in Y_r with a feature in X, $\#S = y_{rj} = x_i$, or the (partial) identity of two features ($\#S = y_j = x_i$), provided that $y_j = y_j + y_{rj}$. The value of $\#S$ thus may be calculated in various ways. It could be the overlap between relations (M. Black): $\#S_{rr} = X_r \cap Y_r$; it could be the overlap between relations in Y_r with normal features in X (Tourangeau & Sternberg): $\#S_{xr} = X \cap Y_r$ (cf. Figure 3.1), or it could be any other combination. These alternatives for calculating $\#S$ in agreement with interaction theory will be discussed in Chapter 5.

In the *response phase*, $\#S_{xr}$ has been assigned a value, and the number of shared (or fitting) relations is compared with the 'threshold' criterion q_r . If q_r is met or exceeded, the expression is read as a metaphor. If $\#S_r$ is smaller than q_r , the expression is an anomaly. Regarding response times, the interaction model does not compete with the comparison and anomaly model. There is a difference in RT between decisions for 'literal' and 'metaphor', whereas there is none between 'metaphor' and 'anomaly'. The latter is absent, since the responses come from the same decision diamond.

A difference between the models is accomplished, only if a variant of the anomaly model is operative (Section 2.3). After the decision 'not literal', the conclusion of 'anomaly' may stop the entire process, without further checks on metaphoricity. In that case, the anomaly model predicts that RTs for (certain) anomalies lie between those for literals and metaphors, whereas the interaction model predicts that RTs for anomalies always equal those for metaphors. Interaction theory does not predict special electrocortical effects for metaphors.

Some summarization of the above may lead to the following formal computation of the two-stage interaction model:

1. encode *A-term*: $X = \{x_1 .....x_n\}$
2. encode *B-term*: $Y = \{y_1 .....y_m\}$
3. if $\#S = X \cap Y \geq q \rightarrow$ respond 'literal' and stop
   else go to 4
4. construct relations for *B-term* features $Y_r = \{y_{r1} .....y_{rm}\}$
5. apply Y_r to X
   if $\#S_{xr} = X \cap Y_r \geq q_r \rightarrow$ respond 'metaphor'
   else respond 'anomaly'
6. stop

q = number of elements in $X \cap Y$ needed to respond 'literal'
q_r = number of relation elements in $X \cap Y_r$ needed to respond 'metaphor'

As already indicated in Section 1.3 (Chapter 1) and 2.3 (Chapter 2), a more subtle version of the model provided in Figure 3.1 uses weighted features and relations. High frequency, prototypicallity, or context cueing, could lead to a preferred order of comparison. The weighted relation set Y_r can be characterized as:

$$W_{Yr} = \sum_{u \, \varepsilon \, Yr} w_{Yr}(u) \; .$$

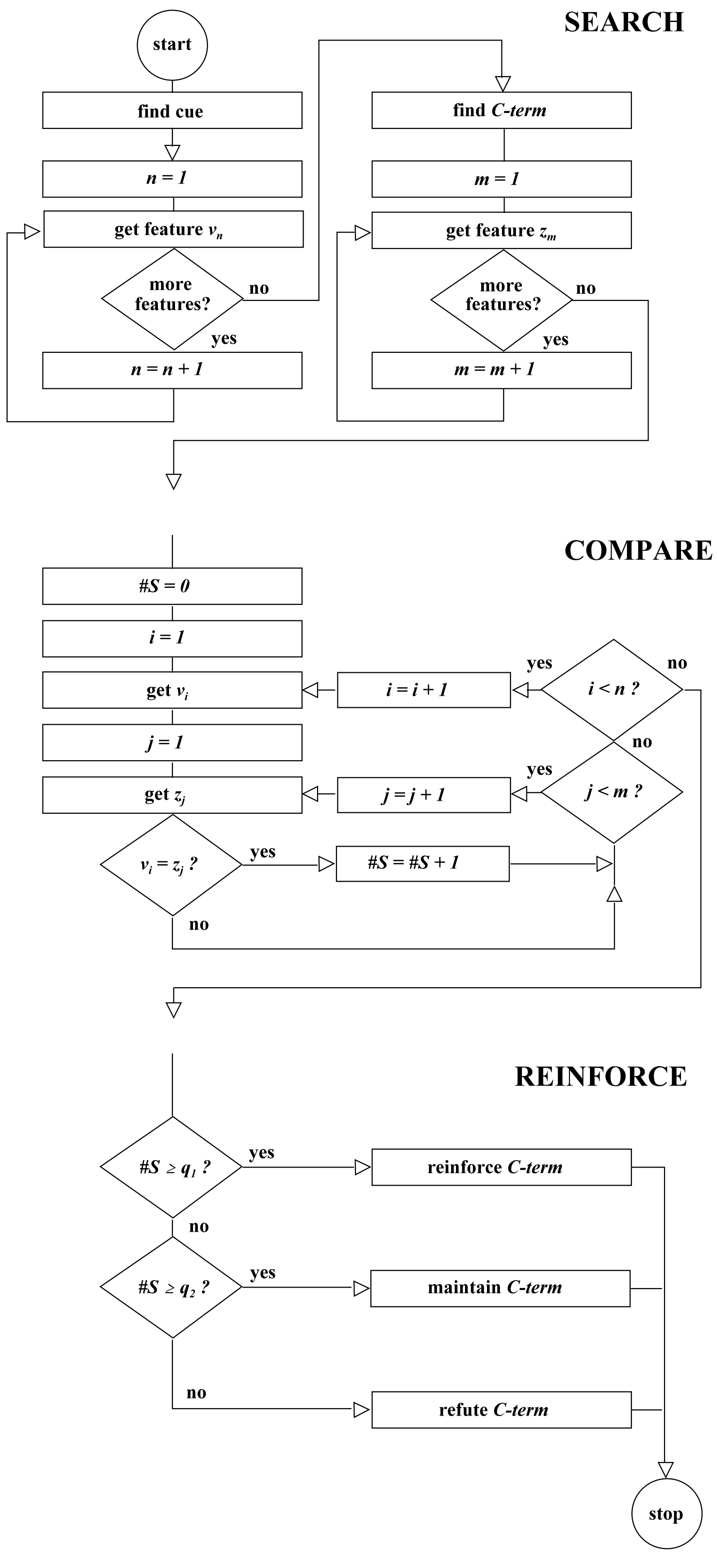
The weighted feature set X is described in Section 1.3. The size of the weighted shared set between normal features in X and relations in Y_r is:

$$W_{Sxr} = \sum_{u \, \varepsilon \, Sxr} w_{Sxr}(u) \; .$$

The interaction theory is the only theory that considers the role of context for the processing of a metaphor. Therefore, a flow-diagram for precues (Figure 3.2) is presented in this Section. In principle, this diagram can also be integrated with the comparison and anomaly model. Yet, precues lead to other predictions in the interaction model exclusively.

Figure 3.2 provides a model for (pre)cues in the creation of the *C-term*. Certain properties of this precue-model are inspired by the ideas

Figure 3.2: Serial precue-model for the calculation of associative power between cue and *C-term*.



that underlie the SAM-model ('search of associative memory') by Raaijmakers & Shiffrin (1980, 1981).

The three metaphor theories suggest that one or more features in the interpretation of the '*A* is *B*' expression should be considered the *C-term*. The comparison theory, for example, states that the interpretation of 'love is a rose' is identical to finding the shared set (for instance, 'red', 'tender', 'thorny'). These features are considered the *C-term*. The other theories device slightly different mechanisms. However, they all claim that there is a feature z in the interpretation set Z that may be considered the *C-term*. The paramount assumption in Figure 3.2 is that the power of association between a cue and the *C-term* depends on the number of matches between the features that the cue activates and the features of which the *C-term* consists.

The process begins with the *search phase*, in which a cue should be found that is useful in the interpretation of the expression. This cue can be anything: Word morphology, phonology, orthography, syntax, semantics, title or author. The cue activates features, forming 'a kind of index for a specific construction, both a set of permissions and a set of prohibitions', as Ricœur (1974: 104) puts it. Figure 3.2 is illustrated by the example of one cue only, activating feature set V. Obviously, the number of cues can be extended infinitely. If there are no features activated by the cue(s) anymore, the permissions and prohibitions for the *C-term* are set. The *C-term* is found in three different ways, according to the comparison, anomaly or interaction view. In the search phase of the precue-model, 'find *C-term*' can be substituted by one of the models in Figure 1.1 (Section 1.3), Figure 2.1 (Section 2.3) and Figure 3.1 (this Section). 'Find *C-term*' should not be read as if the *C-term* itself activates features. However, it is only an inventory of the features in Z that are supposed to form the *C-term* (*C-term* $\subset$ Z).

The associative power of the cue, justifying the appropriateness of the *C-term*, is assumed to depend on the ratio of shared and distinctive features. Thus, the *comparison phase* calculates $\#S$ as the number of shared features between cue (v_i) and the *C-term* (z_j): $v_i = z_j$? . If there is a match, $\#S$ is augmented, if there is a mismatch, a new feature is compared. The loops continue exhaustively, until there are no features left in either set.

In the *reinforcement phase*, the associative power of the cue - expressed by a value for $\#S$ - is checked against the principle q_1 . If $\#S$ is greater or equal to q_1 , the *C-term* will be reinforced. This means that many features of the cue are also found in the *C-term*, so that the latter is strongly supported by the cue. If $\#S$ is less than q_1 , a second criterion q_2 is employed. If $\#S$ is greater or equal to q_2 , the *C-term* will be maintained, which means that the cue is plausible, but not decisive for the interpretation. Further cues will be needed to lend their support. If $\#S$ is less than q_2 , the *C-term* is refuted, which means that the cue lacks associative rigor for the justification of the *C-term*.

1. encode (pre)cue $V = \{v_1.....v_n\}$
2. encode *C-term* $C\text{-}term \subset Z$
3. compute $\#S = V \cap Z$
   if $\#S \geq q_1$ reinforce *C-term*
   else if $\#S \geq q_2$ maintain *C-term*
   else delete *C-term*
4. stop

q_1 = number of elements needed in $V \cap Z$ to reinforce the *C-term*
q_2 = number of elements needed in $V \cap Z$ to maintain the *C-term*

Note that if certain cue-features or aspects of the *C-term* are seen as more important than others, again weights can be employed.

From the models offered in this Section, test predictions for the interaction theory of metaphor processing can be derived. These predictions are partly competitive with the predictions derived from the comparison model (Section 1.3) and the anomaly model (Section 2.3).

1. **A metaphor activates three feature sets**
2. **These three sets need not be fixed**
3. **A literal expression has more shared normal features than a metaphor and an anomaly; a metaphor has more shared (or fitting) relations than a literal and an anomalous expression**
4. **The set of shared features or relations is not fixed and therefore, cannot be predicted from the single terms**
5. **The response time to understand a literal expression is shorter than for metaphors and for metaphors as long as for anomalies**
6. **No special electrocortical effects are expected for metaphors**
7. **Context precues influence 2, 4, and 5**

Outlook on the next Chapters

In the Chapters 1, 2, and 3, the various metaphor theories were analyzed as formal models of computational rules. Chapters 4, 5, 6, and 7 shall be more empirically oriented. Chapter 4 centers upon the question - which metaphors should be selected for the tests on the metaphor models. Apparently, none of the models allows for factors such as word frequency, lexical ambiguity, rhyme, and category verification. Nonetheless, these basic linguistic properties might affect metaphor processing, and perhaps, even explain the differences between expression types. The stimuli thus should be matched on these variables, which form the constraints on the selection.

Whereas Chapter 4 provides pretests on the stimulus materials, Chapter 5 puts the metaphor models to their first test. The activation and comparison of features were captured in the associative flow of subjects who summed up features for the three expression types in three different conditions: For the *single terms*, the *expressions*, and for the expressions in *context*. Moreover, this Chapter examines and amends the concepts of similarity, identity, and salience of features.

Chapter 6 tries to track down the time course of metaphor processing, by that investigating the effects of experimental design on reaction times. In the experiments described in this Chapter, the completion of the *A-term* by the *B-term* made an expression either literal (L), metaphoric (M) or anomalous (A). Subjects were asked to make speeded decisions in a two-choice (L-M, L-A, M-A) and three-choice task (L-M-A). Similar experiments were performed with similes.

The three-choice task exploited in Chapter 6 was replicated in the experiments on the electrocortical effects of metaphor processing in Chapter 7. While subjects performed a reaction time task, the EEG was sampled during the presentation of the *B-term*. The event-related brain potentials (ERPs) derived from the raw EEG were investigated on their sensitivity for processing literal, metaphoric and anomalous expressions. ERP waveforms were supposed to change as a function of expression type. The following Chapter, however, first examines the stability of the stimulus set.

Note:

1. Steen (personal communication) objected to a two-stage model for the interaction view, since features cannot be separated from their relations. They are activated together and are interconnected.

However, the interaction view stressed that what makes a metaphor different from any other expression type is the activation and connection of relations. Thus, literal and anomalous expressions do **not** activate and connect relations as much as metaphors. Ideally (Figure 3.1), they do not at all. Since the creation of relations is a special trait of metaphors, it is unnecessary for processing literal expressions. There is no need to activate relations immediately when a metaphor is encountered, because it should be checked first, whether the expression is literal or not. Else, every metaphor would be recognized as a metaphor in advance, which is circular.

About their interconnectedness, Chapter 5 offers a solution by matching every possible combination of feature type that the interaction theory distinguishes. Subjects wrote down relations behind the features (these features were considered grounds), and the fixed combinations were matched: $X_{rg} \cap Y_{rg}$. However, the result of this analysis was that the shared sets remained empty.

CHAPTER 4: STIMULUS SELECTION[*]

4.0 The research units

> Tricky problems in this field are the methods suitable to give access to knowledge of reading processes (whether on-line or off-line) as well as the samples of texts to be chosen. (Ibsch 1991: 4)

In this Chapter, the stimulus materials are developed to test the metaphor models. What type of metaphor should be chosen? A simple '*A* is *B*'-form or an elaborate Homeric comparison?[1] From what kind of poem? Rhyming or nonrhyming? Basic linguistic properties, such as rhyme, spelling, and number of syllables may affect the processing of a metaphor, and thus, should be kept constant.

In the Sections 4.1 and 4.2, various linguistic properties are discussed which may interfere with metaphor processing. They are seen as constraints a priori on the stimulus selection. Section 4.4 presents tests on the stability of the expression types, while in 4.5, post hoc controls are performed on, for instance, word frequency. Section 4.6 discusses the relevance of this investigation for the theory of literature.

It could be argued that no matter how complex a metaphor is, or in which form it is written, it can always be reduced to a simple '*A* is *B*'-form (Burke 1945: 503-504; Mooij 1976: 26; M. Black 1980: 25; Beardsley 1982: 268). For instance, a genitive construction such as 'a fleece of fog'[2] may be re-written as 'the fog is a fleece'; 'donut shaped accelerators' (Nambu 1985: 126) may be rewritten as 'accelerators are donuts' (where 'shape' is the *C-term*).

It is sensible to understand the processing of a basic metaphor form first, before further factors are engaged. The limits of this study are set, by using metaphors of the form 'the *A* is a *B*',[3] with further restrictions on rhyme, syllables, orthography, word class, function and lexical ambiguity. On the other hand, these metaphors are selected from existing literary texts, so that they are valid for a natural reading situation.

4.1 The theory of literature: Analysis of linguistic properties

Since the work of the Russian Formalists, the theory of literature showed a profound awareness of the text constituting function of linguistic properties. Today, it closely examines the complexity and interrelationships of rhyme, spelling, syllables, syntax and lexical ambiguity in individual works of art. These exhaustive descriptions of the linguistic properties usually serve one purpose only, namely, pointing out how they are responsible for a particular

---

[*] Notes are on the pages 101-114.

interpretation of the work; what their effect is on constituting meaning for the individual.

In this Section, the theory of literature is discussed with respect to formal properties such as spelling, rhyme, syllables, and lexical ambiguity. The sole purpose of this discussion is to make an inventory of linguistic aspects that may influence the processing of a metaphor. To isolate what is unique for metaphor processing, all other aspects of normal reading - such as the influence of rhyme or spelling - ought to be avoided in the experiments. The effects of these linguistic properties are interesting in their own right. However, if they are not kept under control, they are obfuscating what is specific to metaphor processing. Put more technically, the effects of basic linguistic properties could covary with effects of metaphor processing, and thus, should be eliminated. Whereas Section 4.1 is an inventory of what is considered important for constituting poetic meaning, Section 4.2 is a display of what is found out about these linguistic properties in psychology. Section 4.3 considers these findings for the selection of metaphors, literals and anomalies, and provides further tests on their linguistic characteristics, before they are submitted to the experiments in the Chapters 5 up to 7.

Constraint 1: Spelling and rhyme

> An analysis of any linguistic sign whatever can be performed only on condition that its sensible aspect be examined in the light of its intelligible aspect (...) and vice versa. The indissoluble dualism of any linguistic sign is the starting point of present-day linguistics in its stubborn struggle on two fronts. Sound and meaning - both these fields have to be thoroughly incorporated into the science of language: speech sounds must be consistently analyzed in regard to meaning, and meaning, in its turn, must be analyzed with reference to the sound form. (Jakobson 1971: 104)

The practice of literary interpretation usually maintains the firmly rooted principle that where the text provides formal correspondence, semantic correspondence also is intended. Although formal equivalences may establish a close link among semantic units, this does not lead to the conclusion that they always *do* establish them, let alone that they *have to*. Taken as a prescriptive theory on how to interpret, such a formal-semantic equivalence theory may well function. However, taken as a descriptive theory of reader interpretation, it is too meager, and its generalizing power is highly overestimated, as will be shown in Section 4.2.

An outstanding and often quoted advocate of the formal-semantic equivalence view is the above cited Jakobson. He pressed the point that an equivalence on any linguistic level (which may be spelling, rhyme, meter, grammar) leads to a linkage in meaning.

> Briefly, equivalence in sound, projected into the sequence as its constitutive principle, inevitably involves semantic equivalence, and on any linguistic level any constituent of such a sequence prompts one of the two correlative experiences which Hopkins neatly defines as "comparison for likeness' sake" and "comparison for unlikeness' sake". (Jakobson 1981: 40; Hopkins 1966, originally 1865: 106)

Words that rhyme, often look alike. Owing to the fact that the formal-semantic equivalence theory is not limited to rhyme, Lotman (1976) considered its value for orthography:

> In cases where we detect in the graphics an intentional organization we are justified in speaking of their poetic meaning in that everything that is organized in poetry becomes meaningful. (...) It may be assumed that in those cases where the graphic system coincides with the phonological system and they are both present in the mind of the native speaker as a single system, graphics more rarely become a bearer of poetic meaning. But in cases where the automatic character of their association is disrupted and a conflict is felt between these systems, the possibility of imbuing the graphics with poetic meaning arises. (Lotman 1976: 71)

Thus, in normal rhyme situations (e.g., match-patch: Polich, McCarthy, Wang & Donchin 1983), phonology rather than orthography constitutes poetic meaning. When words rhyme, but do not look alike (e.g., blare-stair), orthography may bear poetic meaning as much as phonology. Consequently, in word pairs that graphically match but do not rhyme (e.g., catch-watch), orthography should be strongest in conveying poetic meaning. Word pairs that do not rhyme and do not match orthographically (e.g., shirt-witch) should have the least poetic meaning.

Accordingly, the processes described by the metaphor models may be disrupted by any word in the text (which also may be the *A-term*) that rhymes with, is spelled equivalent to, or is put in a corresponding grammatical position with the *B-term* of the metaphor. In cases where the formal-semantic equivalence theory is operative, formally equivalent terms are enriched with semantic features - or certain semantic features are highlighted - due to the extra correspondence that is created.

Thus, the chance to become metaphoric is higher for an expression with two rhyming terms than for an expression with nonrhyming terms. Apart from the semantic correspondence, a phonological correspondence is established as well for a rhyming expression. Therefore, the number of shared features (semantic plus phonological) is higher for a rhyming expression than for a nonrhyming expression (only semantic), so that the 'threshold' criterion for metaphoricity is more easily met.

Consider also the following. Suppose that a reader encounters an enveloping rhyme pattern in a poem with quatrains. Every fourth sentence ending, thus rhymes with the first sentence ending of each verse. Suppose also that the last word of the fourth sentence is the *B-term* of a metaphor. The similarity between *A*- and *B-term* is based on semantic correspondence - they share semantic features. However, this similarity may be eclipsed by the similarity between the *B-term* in the fourth sentence and the last word of the first sentence. The rhyme, bringing these words together, ensures a set of shared semantic features that is augmented by the shared phonological features.

Furthermore, rhyming could cause an advantage for one of the metaphor theories. As argued above, shared semantic sets of formally equivalent terms may be enlarged by shared formal features (equal phonology, orthography, etc.). Thus, an expression shares more features when the formal-semantic equivalence theory is operative, than when it is not. Due to the accumulation of shared semantic and formal features, the inner 'threshold' criteria are more easily met. In other words, under the parameters of the formal-semantic equivalence theory, anomalies become metaphors, and metaphors become literals, much quicker than outside this theoretical framework. Unintentionally, formal-semantic equivalence theory favors comparison theory, because more expressions will satisfy the inner threshold criterion for literals, so that a second stage - as presumed by the anomaly and interaction model - will hardly ever occur.

Following in Jakobson's wake, Wellek & Warren ticked off another range of formal aspects involved in rhyme, which could also cause unwarranted intrusions on the 'normal' process of understanding metaphors.

> We may ask what is the semantic function of the syllables which rhyme, whether rhyme is in the suffix (character: register), in the roots (drink: think), or in both (passion: fashion). We may ask from what semantic sphere rhyme-words are selected: whether, for example, they belong to one or several linguistic categories (parts of speech, different cases) or groups of objects. We might want to know what is the semantic relation between the words linked by rhyme, whether they belong to the same semantic context as do many of the common doubles (heart: part, tears: fears) or whether they surprise precisely by the association and juxtaposition of completely divergent semantic spheres. (Wellek & Warren 1955: 161-162)

Preferably, although hardly avoidable, effects of rhyme should not be evoked by the stimulus set. Line-end rhymes are probably the most conspicuous, but rhymes in morphological aspects of the words (cf. Wellek & Warren) should be eluded as well. Yet, if rhyme effects are in the stimulus set, the most important thing should be that they are not systematic. Put differently, that they are interrupted by sufficient cases of nonrhyme.

> A fact seldom explicitly brought out (perhaps just because it is taken for granted as obvious) is that the role played by rhyme in the design of verse is determined no less by the distribution of rhymes in a text than by that of its nonrhymes: both positive and negative constraints are operative in poetry. Clearly this has to be so if rhyme, viewed as a paradigmatic relation, is to perform effectively a syntagmatic organizing function. Otherwise, the positive pattern would be subject to obliteration by casual accumulation of structurally irrelevant factors - the signal would tend to be drowned out by the noise. (Abernathy 1967: 1)

Here, Abernathy stated that rhyme (the signal) is not perceived as such, if its distribution is not systematic; if it is interrupted by accumulative nonrhymes (drowned out by the noise). If this is correct, the effect of nonrhymes is extremely useful to create an unbiased stimulus set. By selecting metaphors from free verse, rhyme may not form the signal in an expression, but spotting its metaphoricity may.

In this connection, Hoorn (1996) conducted a pilot experiment on the electrocortical effects of rhymes and nonrhymes on semantics, while subjects read alternating rhyming verses. It was found that systematic expectations of rhyme interacted strongly with semantics, particularly when these expectations were frustrated by nonrhymes.

The solution to the rhyme problem should be that subjects do not anticipate systematic rhymes. Therefore, **the first constraint on the stimulus selection is that the effects of phonology and orthography are reduced to random noise by selecting nonrhyming metaphors from nonrhyming poems (i.e. free verse)**.

Constraint 2: Syllables and letters

> Syllables are converted into units of measure, and so are morae or stresses. (Jakobson 1981: 27)

Words with more than one syllable will differ in the stress patterns within the word boundary, which is said to be important for the poetic function of the word: 'In poetry one syllable is equalized with any other syllable of the same sequence; word stress is assumed to equal word stress, as unstress equals unstress' (Jakobson 1981: 27).

Syllable patterns may be seen as more than mere outer form. Lotman (1976) discussed the semantic importance of syllable structure in Tyutchev's poem 'Last love':

> Let us look first at the problem of isometrism insofar as in Lomonosov's syllabo-tonic system, metrical ordering also specifies the recurrence of the number of syllables in a line.
>
> The first stanza affords a certain inertia of expectation by creating the correct alternation of the number of syllables: 8-10-8-10. True, there is already an anomaly here: habit in reading Russian iambic tetrameter, the most widespread meter in post-Pushkin poetry, disposes one to expect the correlation 8-9-8-9. The extra syllable in the even lines is distinctly heard by the ear. Thus, against the background of preceding poetic tradition, the first stanza is a violation. But from the point of immanent structure, it is ideally ordered, and this forces us subsequently to expect precisely this type of alternation. (Lotman 1976: 48)

> (...) each of the two combined types of lines, counter to our expectation, is lengthened by a syllable that is also extremely noticeable to the ear. The 10 syllable line is replaced by the 11 syllable, and the 8 syllable line by the 9 syllable. Here an additional variation is introduced: in the long lines the syllable is lengthened in the first of them (the second is perceived as *shortened*), and in the short lines-in the first (the second is perceived as *lengthened*).
>
> The diverse violations of the established order in the second stanza require in the third stanza a resurrection of the inertia of expectations: the inertia 8-10-8-10 is reintroduced. But here too there is a disruption: in the second line instead of the expected 10 syllable unit, there are 9 syllables. (Lotman 1976: 48)

> In the syllabo-tonic line the disposition of stresses is just as important a factor as the number of syllables. (Lotman 1976: 49)

These quotes from Lotman illustrate that expectations of syllabic patterns (and thus stress patterns) are important to the skilled poetry reader. Certain violations may be disturbing or surprising. When a *B-term* occupies a position in the meter that expects more syllables than the *B-term* provides, a disruption and an effect of surprise may hinder the normal processing of the metaphor.

Thus, for a fair test on the metaphor models, the selected expressions should not be part of poems that provide expectations about the syllable pattern or contain predictable meter. Analogous to rhyme, syllable effects should be reduced to random noise. 'Frustrated expectations' (Jakobson 1981: 33) at the wrong level of processing (i.e. other than those the metaphor models hinge on) ought to be avoided.

> In some patterns of versification the syllable is the only constant unit of verse measure, and a grammatical limit is the only constant line of

> demarcation between measured sequences, whereas in other patterns syllables in turn are dichotomized into more and less prominent, and/or two levels of grammatical limits are distinguished in their metrical function, word boundaries and syntactic pauses.
>
> Except the varieties of the so-called vers libre that are based on conjugate intonations and pauses only, any meter uses the syllable as a unit of measure at least in certain Sections of the verse. (Jakobson 1981: 29)

Resuming, *free verse* seems to be the proper kind of poetry to select the metaphors from, because it avoids the use of syllables as a unit of measure for meter, and discourages systematic rhyme expectations.

What is said for the number of syllables in a word is closely related to the number of letters. With the increase of syllables, the number of letters increases as well. However, more or fewer letters within the syllable may have different effects; they may become more or less prominent:

> In quantitative ("chronemic") verse, long and short syllables are mutually opposed as more and less prominent. This contrast is usually carried out by syllable nuclei, phonemically long and short. (Jakobson 1981: 30)

It may make a difference, whether we read 'love is a rose' or 'love is virtuousness'. Although the first expression may be judged as a metaphor, and the second as a literal expression, the latter could be more 'foregrounding' (more salient), simply because it is a longer word. In terms of psychological measurement, processing the latter expression may take more time than the first, because longer words are more time consuming than shorter words (see Section 4.2). A longer processing time for 'love is virtuousness' is not expected by the metaphor models, and may be caused by the larger number of letters and syllables for 'virtuousness'. In other words, the *B-terms*, which form the target stimuli of the experiments, should consist of equal numbers of syllables and letters.

Regarding context, choosing expressions from free verses, again, is the best solution to avoid expectations about syllable number and word length, because it might just be that 'prosodic long is matched with long, and short with short; word boundary equals word boundary, no boundary equals no boundary' (Jakobson 1981: 27). Thus, **the second constraint on the stimulus selection is that particularly the words that make an expression literal, metaphoric, or anomalous (the *B-terms*) should have equal numbers of syllables and letters**.

Constraint 3: Word frequency

Although a major field in (psycho)linguistic research, word frequency is never considered important in the theory of literature. The main focus on 'meaningful' aspects of language made word frequency a subordinate research object, a 'minor linguistic technicallity'. Only an inferential detour may bring to light that the theory of literature indirectly does speculate about effects of word frequency.

In the theory of foregrounding (cf. Šklovskij 1965), automatized language is transgressed by salient, aberrant words. Assumingly, normal words have less importance for the poetic meaning than unusual ones. Within this framework, word frequency differences could be the explaining factor behind this effect. Ordinary language tends to use high-frequent words, familiar to most language users. If such language is disturbed by interjecting a 'strange', 'unusual', element, this word is an *infrequent* word for that sociolect or for that context. It may even be argued, that theories of surprise, salience, or foregrounding are implicit theories on frequency of use in a particular context. Anything atypical, novel, peculiar, rare, singular, uncommon, or unique, is then a measure of the low frequency of occurrence subtracted by the high-frequency of occurrence of surrounding phenomena.

The first half of a sentence primes the meaning of the second half. Unexpectedness induced by an inadequate sentence ending (e.g., a deviant word), thus may be a function of the low-frequent use of that word in the context of that particular sentence.

To tackle the problem from the opposite angle, repetition of a deviant word (or the reintroduction of a character in a novel) may take away all of its 'strangeness'. Since repetition augments the frequency of use by 1, each time the word is repeated (or the character presented), its strangeness is attenuated accordingly. Taking into account Lotman's (1972: 125) statement that 'repetition is equally important as equivalence', word frequency as an equivalent of repetition probably is fundamental to language perception, including the perception of literature.

In a way, word frequency is relative to the position in the text. Three factors may be distinguished in this respect. Firstly, the position of the word in the text, either with many, few, or no preceding words. Secondly, the high or low frequency of the word in the speech community. Thirdly, the sort of priming, which may be correct or incorrect. If the strange word is in the first position of the text, no priming occurs, and a potential surprise effect may be fully ascribed to the infrequency of the word. If the strange word is not in first position, the priming effect increases with the number of words that precedes it. The effect of infrequency becomes higher, accordingly as the prime is less correct. The effect of infrequency decreases in proportion to the correctness of the prime. Rereading the text, and then interpreting the strange word in the first position, is the same (in the sense of a 'backward prime') as interpreting a word with many preceding words, enhancing the effect of priming.

The further a word is positioned, the higher the priming effect of the preceding words (either correct or incorrect). A correct prime of a high-frequent word renders no surprise (ordinary language). A correct prime of a low-frequent word also renders no surprise, or perhaps a little, due to the infrequency. An incorrect prime of a high-frequent word does render a surprise effect, due to the prime (effect of context). An incorrect prime of a low-frequent word renders the largest surprise, because both factors add, or interact.

B-terms are primed by the *A-term*, and by the surrounding context. In résumé, **the third constraint is that *B-terms* ought to be in the same word frequency bandwidth, to avoid word frequency effects relative to expression type**. Metaphoricity, literalness, or anomality should not be explained by systematic word frequency differences among the expression types, but by their presumed special processes.

Constraint 4: Lexical ambiguity (homonymy)

Homonyms are the mimicry of meaning. When a word takes on one meaning, it is not the other; when it takes on the other meaning, it is not the first. A simple example is the word 'bank', which lies alongside a river, or is an establishment for keeping money. Hardly ever does context allow all meanings to be active (cf. Section 4.2), unless there is no context available.

> For the receiver the message presents many ambiguities which were unequivocal for the sender. The ambiguities of pun and poetry utilize this input property for the output. (Jakobson 1961: 249)

> Ambiguity is an intrinsic, inalienable character of any self-focused message, briefly, a corollary feature of poetry. Let us repeat with Empson: "The machinations of ambiguity are among the very roots of poetry". (Jakobson 1981: 42)

The problem with lexical ambiguity in the study of metaphor is the doubt of the reader. In choosing for one of the meanings of an ambiguous *B-term*, the continuation of the process may be impeded in comparison with an unambiguous *B-term*. On the other hand, when the *B-term* is ambiguous, chances improve for expressions to become metaphoric, because more candidates are available for establishing figurative meaning. According to **the fourth constraint, ideally no lexical ambiguities are allowed to enter the experimental expression set**. However, in natural language and specifically in poetry, lexical ambiguity is hardly avoidable. Expressions containing lexical ambiguities are listed in Table 4.3. After the experiments have been run, these expressions are checked for deviant results.

Constraint 5: Linguistic categories

With words the world is described, with words categories are formed to describe the world by. Greimas put this nicely, when he stated:

> (...) les classèmes (...) rendait compte de la linéarité sémantique, relativement homogène, du discours. (...) un nouveau rôle qu'on peut attribuer aux classèmes (...) est de constituer le cadre de l'organisation de l'univers sémantique. (...) C'est ici qu'apparaît une nouvelle fonction, classificatoire, d'une certaine espèce de sèmes. (Greimas 1966: 78-79)

People use semantic categories to structure their world. These categories are not formed randomly. In fact, Section 4.2 reviews research that shows that people use categories according to certain psychological principles. At least at the theoretical level, it must be understood how metaphor processing may be intertwined with category verification. A metaphor that combines an instance as the *A-term* with a category as the *B-term* (e.g., 'love is botany') may be processed differently from a metaphor that combines a category as the *A-term* with an instance as the *B-term* (e.g., 'emotion is a rose'), just because the instance-category distribution differs. In a semiotic study of language, Greimas proposed how to systematize the species-genera (instance-category) relationship between words, which will be used in this Section to form a basic understanding of what the problematic is about.

> On voit que les classèmes
>
> objet vs animal vs humain
>
> semblent pouvoir être articulés en un système sémique, qui serait peut-être plus explicite si on le présentait ainsi:
>
> inanimé vs animé
>
> animal vs humain
>
> La généralisation que nous voulons proposer dès maintenant serait la suivante: si les figures sémiques, simples ou complexes, relèvent du niveau sémiologique global, dont elles ne sont que des articulations particulières prêtes à s'investir dans le discours, les classèmes, de leur

> côté, se constituent en systèmes de caractère différent, et appartiennent au niveau sémantique global, dont la manifestation garantit l'isotopie des messages et des textes. (Greimas 1966: 54)

First, Greimas illustrated that objects can be ordered hierachically. There are higher-order categories, such as 'inanimate' vs 'animate', which can be subdivided in intermediate categories ('animate' e.g., contains the intermediate categories 'animal' and 'human'). It is easily understood that these intermediate categories themselves can be subdivided into lower-order categories, such as 'monkeys' and 'horses' for the animals, and 'man' and 'women' for the humans. Of course, these subdivisions can go on infinitely, resulting in all species (objects) that belong to the category.

In set theoretical terms, a category is a set of elements that share the category-specific features, i.e. those features not shared by any other category. The frequency of category-specific features shared by all members is higher than the frequency of those features in other categories. This may be thought of, when Greimas mentioned 'the isotope of messages and texts'. The next paragraph will explain the idea.

Greimas posited that when a category-word is encountered, appropriate instances of the category are evoked - this is the paradigm - which may be affirmed by the text with relevant instances. Thus, in the text, selection from the paradigm (here, the category members) created a 'linked chain of associates', which is called an isotope. An isotope, then, is a set of features (associates) connected by, in this case, category membership. In other words, when people read a category-word, it is likely that they associate instances with category-specific features, thus forming a paradigm. Reversely, when people read an instance-word, it is likely that they associate the category-word. Thus, the frequency of mentioning these words in response to each other will be high. Henceforth, the likelihood that an instance is associated, given the category word is called the 'instance dominance' of a word. The likelyhood that a category is associated, given the instance is called the 'category dominance' of a word. Chapter 5 will draw on this matter again.

The **fifth constraint states that expression types should not differ in the way instances and categories are distributed over *A*- and *B-term***. In Section 4.5 and 4.6, however, it is argued that literals, metaphors and anomalies show systematic differences in the way in which categories are distributed. Therefore, category verification may systematically differ among expression types, thus forming an alternative explanation for the metaphor models.

Constraint 6: Syntax

Analogous to rhyme, grammar may also function according to the formal-semantic equivalence assumption, in that 'syntactic pause equals syntactic pause, no pause equals no pause' (Jakobson 1981: 27). Parallel syntactic structures in a poem could bear semantic parallels as well.

Jakobson stated that rhyming words link up more closely when they share grammatical aspects, such as word class and syntactic function. Considering metaphor processing, *A*- and *B-terms* may be liable to the same effect.

> Rhyme is usually defined as correspondence in terminal sounds, but at the same time it always matters whether the rhyming elements are merely homophonous or whether they are grammatically identical - whether the rhyme links identical formal units or different formal units but belonging to words of one and the same word-class. Do the rhyming words have similar or different syntactic functions? The rhyme technique of diverse poets and poetic schools can be grammatical or antigrammatical, but it cannot be agrammatical. This means that the relation between the phonemic and grammatical structure of rhyme always remains pertinent. (...) Both rhyme and grammatical parallelism necessarily and simultaneously present both of these aspects, but with the difference that in rhyme the emphasis is on the phonemic structure and in parallelism the predominant role belongs to the grammatical aspect. (Jakobson 1971: 111)

What may be valid for rhyme, may also apply to metaphors. The similarity between *A*- and *B-term* may be upgraded when equal word classes are used. The *A*- and *B-term* in 'war is a sword' combine two nouns and may be judged as more equivalent than in a combination of a personal pronoun and a noun ('she is a sword'), or a proper name and a noun ('Anna is a sword').

Likewise, the more the syntactic functions in *A*- and *B-term* differ, the more their degree of similarity may differ. 'War' and 'sword' maintain a subject-nomalized predicate relation in 'war is a sword', and the *A*- and *B-term* may be said to be functionally in balance. When e.g., an adjunct is added ('war is a sword of ice'), then the *B-term* has a nomalized predicate, containing a prepositional phrase, while the *A-term* still has the one function of subject. Similarity ratings may now be different from 'war is a sword', not only because more semantical information is incorporated in the *B-term*, but also as a consequence of more syntactical functions in the *B-term*. In other words, when a term is semantically upgraded, it automatically receives more syntactical features as well.

Similarity in syntactic function may be restored, when the *A-term* also receives a prepositional phrase, as in 'beauty as weapon is a sword of ice'. It may be, then, that 'beauty' is matched against 'sword', as 'weapon' is against 'ice', simply because of the functional equivalence. When another prepositional phrase is added in the *A-term*, as in 'beauty as weapon against war is a sword of ice',[4] it may even be argued that the imbalance of the two prepositional phrases in the *A-term* ('as weapon', and 'against war'), and the one in the *B-term* ('of ice'), overpowers the balance between the subject nucleus ('beauty') and predicate nucleus ('sword'). Matching the secundary syntactic functions (prepositional phrases) may become more important than the primary functions of subject nucleus and predicate nucleus, thereby changing the focus of the comparison from the semantics of the words in the primary functions, to those occupying the secundary functions.

The **sixth constraint, then, proclaims that expressions included in the experiments should not differ in word class and grammatical function for the *A*- and *B-terms*. Tense and person of the verb also should be equal for all expressions**. To achieve a high syntactical consistency within and between each stimulus group (literals, metaphors, anomalies), *A*- and *B-terms* will be singular nouns in present tense sentences of the form 'the *A* is a *B*'. In Section 4.2, it is discussed that the *A*- and *B-term* distribution does not follow the subject-first, nomalized predicate-last order per se. In Section 4.4, a test is provided to control for that.

Summarizing Section 4.1, it could be argued that understanding a metaphor in a literary text may involve more than finding similarity between *A*- and *B-term*, changing from literal to figurative meaning, or creating relations, as suggested by the metaphor models. Other linguistic aspects could influence the assumed processes. Phonological anticipations, orthographical (mis)matches, category verification, lexical ambiguity, and word frequency may disturb the expected processes. The basic linguistic properties should be kept under control by selecting expressions checked or tested on all these dimensions. Put differently, the linguistic properties should not become covariates (should not explain the differences) of processing different expression types.

In the next Section, studies on the cognitive effects of the linguistic properties are discussed. As to the question which expression should be considered a metaphor, and which elements can be regarded as *A*- or *B-term*, a test is provided in Section 4.4 on the categorization of the expressions, and *A*- and *B-term* distribution.

4.2 Psychology: Effects of linguistic properties

Processing a metaphor can be seen as a 'higher-order' cognitive activity. Therefore, it is important to know the extent to which linguistic properties - i.e. the sentence characteristics on word level - contribute to the processes suggested by the three metaphor models. In this Section, psychological studies on linguistic properties will be discussed. The results will be used for the selection of stimulus materials, on the dimensions of orthography, phonology, syntax, and word frequency. The actual test on this selection, and on the *A*- and *B-term* distribution within the expressions, will be performed in the categorization experiment, described in Section 4.3.

Controlling lexical ambiguity (homonymy), instance-category distribution, instance-category dominance, and again, word frequency will be post hoc, because uniformity on these criteria beforehand, left too small a stimulus set. Moreover, instance-category dominance can only be determined after a feature elicitation experiment. The above criteria will serve as a check list to account for possible covariances in the feature elicitation, RT and EEG experiments.

Constraint 1: No expectations for spelling and rhyme

Reading rhyme in writing is not only the matching of words in a pure phonological comparison, but also an implicit spelling or typographical check in an orthographical comparison. Consequently, a match on the phonological level, supported by a match on the orthographical level, could reinforce the perceived similarity between the *A*- and *B-term* (cf. Jakobson 1981: 40).

The mingling of phonology and orthography may also be illustrated by the findings of orthographic effects on rhyme monitoring, and of phonological effects on visual word recognition. Seidenberg & Tanenhaus (1979) reported that subjects showed no awareness of having accessed orthographic information in a rhyming task. Reversely, Meyer, Schvaneveldt & Ruddy (1975) showed that phonological information also was activated in visual word recognition. In a Stroop paradigm used by Tanenhaus, Flanigan & Seidenberg (1980) the same results were obtained for color naming latencies.

Polich, McCarthy, Wang & Donchin (1983) and Kramer & Donchin (1987) also found interactions between the effects of orthography and phonology. In their studies, subjects compared visually presented word pairs (e.g., match-patch, blare-stair, catch-watch, shirt-witch) for similarity on orthographical and phonological dimensions. Subjects were asked to make rapid decisions (match/no-match?), when word pairs were orthographically and/or phonologically same or different. Polich et al. and Kramer et al. found in RT as well as in the recorded EEG, that subjects were able to identify the orthographical and phonological characteristics of a word within 260 ms. Analyses of RT and EEG showed that readers were not able to skip the orthographical analysis when asked to ignore the orthographical differences, but that they could skip the phonological analysis, when asked to ignore phonological differences. Differences of one letter did not show any significant effect on RT or EEG. Additional evidence for phonological and orthographical interactions in the EEG was found by Rugg & Barrett (1987).

Rugg (1984) in a study on rhyme judgement, and Hoorn (1996) in a study on interactions between phonology and semantics in reading small verses, showed that one component in the EEG (the event related brain potential N400) is also influenced by phonological mismatches, although it was supposed to be sensitive to semantic mismatches only. Negative shifts in the EEG - similar to N400 - were also found for phonological mismatches by Sanquist, Rohrbaugh, Syndulko & Lindsley (1980). Rugg & Barrett (1987) discriminated between a unilateral, right hemispheric, rhyme sensitive N450, and a more bilateral, semantically sensitive N400, but they stressed that these components can also be regarded as one, influenced by different sources of information.

In short, orthographical encoding interacts with phonological encoding, and both interact with semantic processing. For a fair test of the metaphor models, **the first constraint on the stimulus selection is that rhyme and spelling anticipations for the *B-term*, based on the *A-term* or surrounding context, ought to be eliminated**. As suggested in Section 4.1, this is best achieved by choosing expressions from poems written in *free verse,* which have no predictable rhyme or spelling patterns.

Constraint 2: Small numbers of syllables and letters

Flesh (1948) presented a statistical formula for the measurement of readability (comprehension difficulty), based on the number of words per sentence and the number of syllables per word, the Reading Ease formula (R.E.):

Systematically select 100-word samples from the material to be rated
Determine the number of syllables per 100 words (w_l)
Determine the average number of words per sentence (s_l)
Apply in equation R.E. = $206.835 - .846w_l - 1.015\ s_l$

According to Flesh, the measurement of word length is an indirect measurement of word complexity, which is, in turn, an indirect measurement of abstraction. The more syllables the word has, the more complex the abstraction level of the word is.[5] Stevens & Stone (1947) demonstrated the predictive power of this formula, yielding scores that correlated .7 with subject ratings. Klare (1963: Table 1, 76-80) showed a summary of readability formulas, all focused on average sentence length in words, and average word length in syllables. Differences of one letter in word length usually have no effect on readability (Polich, McCarthy, Wang & Donchin 1983; Kramer & Donchin 1987).

Number of syllables within a word, and number of words within a sentence can influence readability. Therefore, **the second constraint on stimulus selection demands that the number of letters and syllables in the *B-term*, and the number of words in the expression are equal**.

Constraint 3: Equal word frequency bandwidths

Literary metaphors consist of words. Two theories of metaphor processing, comparison theory and anomaly theory, claim that the *C-term* of a metaphor can be predicted from the feature sets of the single words. If this is valid, familiarity of the word is a relevant variable. Since familiar words usually describe familiar objects, they activate the largest feature sets. In turn, word frequency is an aspect of familiarity. Familiar objects are often encountered, so that the corresponding words are frequently used. As a rule of thumb, words frequently used in a speech community refer to objects (or concepts) that are highly familiar to that community.

If the *C-term* of a metaphor can be derived from the isolated words, word frequency might have an effect on the probability that the *C-term* will be established at all. When a metaphor consists of two low-frequent words, the chance that a *C-term* is formed may be reduced, because a smaller number of features can be matched, or a smaller number of relations can be created. In contrast, the probability of forming a *C-term* increases, when two high-frequent words are used in the metaphor. As a consequence, word frequency should be reflected in the feature elicitation rates and expression judgement time. Note, however, that context may affect the identification of low-frequent words and the elicitation of features.

The classic word frequency effect is that high-frequent words are recognized faster than low-frequent words (Solomon & Postman 1952; Postman & Conger 1954; Morton 1969; Rubenstein, Garfield & Millikan 1970; Rubenstein, Lewis & Rubenstein 1971). Likewise, high-frequent meanings of lexical ambiguities are recognized faster than low-frequent meanings (Simpson 1981; Jastrzembski 1981). Furthermore, high- and low-frequent meanings of lexical ambiguities evoke different electrocortical effects (Van Petten & Kutas 1987).

However, a number of semantic variables overrule this straightforward effect. Jastrzembski (1981) showed that a low-frequent word with multiple meanings (or with a large cluster of meanings for an etymological derivation of the word), is recognized faster than a high-frequent word with only one meaning (or a small cluster for a derivation). Number of meanings or number of related meanings of a word, therefore, is more important to lexical access than mere word frequency.

Moreover, tasks with high semantic loading also reduce the word frequency effect. Balota & Chumbley (1984) showed that word frequency had no influence on RT in a category verification task, which is supposed to involve lexical access. Subjects matched a category name (e.g., bird) with an exemplar of that category (e.g., robin), or with an exemplar of a different category (e.g., sofa). Instance dominance (likelihood of producing the exemplar given the category name) and category dominance (availability of the category name given the exemplar) had strong effects on verification time. However, there was no effect of word frequency on either correct yes- or correct no-responses.[6]

Another factor that overrules the frequency effect is semantic priming. Balota & Chumbley (1984) primed instances by their correct category, which reduced the word frequency effect in correct yes-responses. Low-frequent words were processed faster with an appropriate prime. Earlier evidence for the effect of semantic priming on word frequency and lexical access was found by James (1975), Becker (1979, 1980), and Whaley (1978). However, Balota & Chumbley noticed that also correct no-responses to inappropriate primes did not show word frequency effects. In other words, correct verifications were insensitive to word frequency differences. Further evidence was provided by Millward, Rice & Corbett (1975), finding word frequency effects, neither for instance-category nor for category-instance correct no-decisions. Balota & Chumbley argued that these results were not predicted by older lexical access models. If in a verification task, a misprimed category evokes a correct no-response, these models expect word frequency to independently affect response time, because the verification requires lexical access.

In other words, subjects are able to make a 'short cut', deciding that if the target (either an instance or a category) is not in the feature set of the prime (category or instance), the expression is incorrect. Applied to metaphor processing, subjects may not have to activate the *B-term* features, if the *B-term* is not a feature of the *A-term* set. For anomalies, low frequency of the *B-terms* possibly has no effect on response times for correct 'anomaly' decisions, because the *B-term* is not in the *A-term* set, so that further processing can be cancelled. If so, the variant of the anomaly model becomes operative, suggesting that after 'not literal' decisions, anomalies are not necessarily passing through the figurative stage of processing.

Lexical decision tasks (is the stimulus a word or not?) involve single word recognition, without being primed by a context. In using a lexical decision task, Balota & Chumbley elicited the classic finding, that word frequency has a large effect on RT. On the other hand, category dominance and instance dominance of the word also influenced RT in the lexical decision task, and in fact, independently from variation in word frequency. They argued that this is odd, because lexical access models assume that no semantic processing takes place before lexical access. Balota & Chumbley suggested that the lexical decision task may be the wrong tool for studying lexical access, because semantic processing clearly influences the effect of word frequency on lexical access.

The conclusion as to the effect of word frequency on lexical access is that there is an effect on the single word level. However, this effect may be overruled by context priming, the number of meanings or related meanings of a word, category dominance or instance dominance. In the feature elicitation experiment described in Chapter 5, a *single term* condition is employed, which may yield smaller feature sets for low-frequent words. Therefore, **the third constraint on stimulus selection is that particularly the *B-terms* should have equal word frequency bandwidths**. Section 4.5 shows that this is a problem for anomalous *B-terms*.

Constraint 4: Priming dissolves lexical ambiguity (homonymy)

Almost any word has multiple meanings, or at least activates more than one related meaning, and often, ambiguity is intended in metaphors or poetry. If response times slow down by unscrambling lexical ambiguities (e.g., 'star' in 'the sun is a star'), this could interfere with the processes of interest. The point is that the metaphor theories have different assumptions on the encoding phase. Word encoding starts with lexical access ('find *A-term*', 'find *B-term*'), which is followed by feature activation.

As mentioned, comparison and anomaly theory claim fixed feature sets. Encoding the terms remains unaffected by prior knowledge, comprehension of the context, or any other concept-driven analysis. In contradiction with the findings of Balota & Chumbley, variation in lexical access of the *B-term* as a function of (instance or category) dominance in the *A-term* set is not expected. Feature activation in the comparison and anomaly view implies automatic processing in the encoding phase, which is in line with theories of lexical access, put forward by Posner & Snyder (1975), Posner (1978), Shiffrin & Schneider (1977), Hasher & Zacks (1979), and Forster (1979).

Contrariwise, interaction theory insists on the creativity of the reader, the bias of context, the individual's world knowledge and controlled processing, which implies a less data-driven analysis. These ideas are more in line with the results reported by Meyer, Schvaneveldt & Ruddy (1975), Rumelhart & Ortony (1977), and Fischler & Bloom (1979). Feature activation in the interaction view implies nonautonomous lexical access, which means that the multiple meanings of, for instance, lexical ambiguities are solved by the context, while in the comparison and anomaly view, multiple meanings would survive.

In the lexical ambiguity literature, there is evidence for both automatic activation and controlled selection of meaning. The crucial question seems to be which moment in the encoding phase is broached by the experiment, i.e. the moment of measurement defined by the inter-stimulus-interval between a contextual prime and an ambiguous target. The initial encoding of a word on the orthographical and phonological level, and the access of multiple meanings, take place between about 150-300 ms. after word presentation (Seidenberg, Tanenhaus, Leiman & Bienkowski 1982; Polich, McCarthy, Wang & Donchin 1983; Kramer & Donchin 1987; Till, Mross & Kintsch 1988). Seidenberg, Tanenhaus, Leiman & Bienkowski (1982) reported six experiments, in which subjects responded almost immediately (within 200 ms) to both meanings of an ambiguity. For ambiguities consisting of two nouns, or a noun and a verb as their competitive meaning, this result was found in various conditions, among which were priming contexts. In other words, in the early encoding of the word, subjects did not use context primes, world or language knowledge, to select a meaning from a lexical ambiguity, thus, showing almost immediate multiple access of meanings. The only exception - selective access - was found by Seidenberg et al. (1982) for noun-noun ambiguities in strong priming contexts. However, the authors did not explain this result in terms of top-down analysis, but as a result of different activation levels, because the nonpriming context condition still yielded multiple access. Context-independent access of multiple meanings in sentences was also found by Swinney (1979), Tanenhaus, Leiman & Seidenberg (1979), Onifer & Swinney (1981) and Kintsch & Mross (1985). Lexical ambiguities embedded in sentences facilitated lexical decisions for related targets, even if these targets were related to meanings that were different from those implied by the sentence context.

Evidence for meaning selection as a function of context priming on longer latencies (500 ms.) in the encoding phase, was also reported by Seidenberg, Tanenhaus, Leiman & Bienkowski (1982), using, among others, noun-noun ambiguities in contextually appropriate and inappropriate contexts. At short latencies both meanings were accessed, independent of the type of context prime, whereas at longer latencies only the contextually appropriate meaning was facilitated. Further support for these results is found in Glucksberg, Kreuz & Rho (1986) and Schvaneveldt, Meyer & Becker (1976).

Reasoning from these data, the comparison and anomaly model could be said to describe the initial, early, encoding of words. The assumption of fixed feature sets is supported by the finding that multiple meanings and related meanings are automatically accessed. It seems an overlearned operation, perhaps anchored in physiology. The interaction model, on the other hand, could be said to describe the final part of this operation: A rapid selection, which occurs at the end of the encoding, sensitive to contextual information and other concept-driven analysis.

With respect to the stimulus materials, **the fourth constraint declares that lexical ambiguities should be excluded from the set**. Lexical ambiguities in the stimulus materials may remain active in *single term* conditions, and in conditions of weak (metaphors) or inappropriate (anomalies) priming. The only disambiguating contexts are literals in *expression* conditions. In Section 4.5 which of the selected stimuli might suffer from lexical ambiguity will be discussed.

Constraint 5: Category verification could intermingle

Metaphors consist of words, and words establish elementary categories with which people structure their world (Rosch 1973; Rosch & Mervis 1975; Rosch, Mervis, Gray, Johnson & Boyes-Braem 1976). Boddy & Weinberg (1981) reported that response latencies are shorter and brain potentials (N1-P2) larger for responses to positive than to negative instances of primed categories. Polich (1985) reported negative shifts (N400) for instance-category mismatches in comparison with matches.

Literal expressions, such as 'the monkey is an animal', might include category verification, matching the positive instance 'monkey' with the category 'animal'. In 'the monkey is a cauliflower' the incorrect instance 'monkey' is applied to the category 'vegetables', but the expression might also be judged as a metaphor. If metaphor processing can be explained by category verification and by finding or creating similarity, how then, are these dimensions related? How homogeneous should the stimulus groups be in terms of instance-category distribution or instance-category dominance?

Certain aspects of category verification, such as dominance, appropriateness, and category hierarchy, have an effect on sentence processing. As stated earlier, instance dominance and category dominance influence the processing time of category verification (Balota & Chumbley 1984). Both phenomena can be explained in terms of feature elicitation.

Instance dominance is defined as the likelihood of producing the exemplar given the category name, and category dominance as the availability of the category name given the exemplar (Balota & Chumbley 1984). Thus, if a literal expression combines a nondominant instance with a nondominant category, the processing of this expression could take much longer than a metaphor that uses a dominant instance and a dominant category. This can be controlled by searching the feature set of the *A-term* for the presence of the *B-term*, and searching the feature set of the *B-term* for the presence of the *A-term*. The relative frequency of mentioning will express the power of dominance, and indicates their mutual associability. In this respect, it can be decided whether the stimulus groups are homogeneous. Chapter 5 describes the results of such an analysis.

Other aspects, such as category hierarchy, could also influence the response to an expression. Higher-order categories, such as 'furniture', have instances ('chair' and 'table'), which themselves could be seen as intermediate categories for very specific instances, such as 'rocking chair' or 'kitchen table'. The latter could be viewed as the lowest category. Rosch

(1973), Rosch & Mervis (1975) and Rosch, Mervis, Gray, Johnson & Boyes-Braem (1976) showed ample support that the intermediate categories are preferred to the higher- and lower-order categories for the semantic organization of the world. The higher-order categories are too abstract, and the lower-order categories too specific, to be frequently used in daily situations.

Reasoning from these results, preference to the intermediate categories may increase their frequency of use. Furthermore, the repetitive use of these categories may influence the degree of familiarity. Repeating a word - in this case a category name - has shown to increase its familiarity and meaningfulness (FM-value), which in turn, decreases recognition speed (Scarborough, Gerard & Cortese 1979). It is stressed by Balota & Chumbley (1984) that higher-order semantic processing is influenced by the familiarity and meaningfulness of a word. As this FM-value is increased, RT is speeded, irrespective of the initial low word frequency. In contrast, RT slows down when the FM-value is lower, despite a potential high word frequency. The FM-value can be increased by contextual priming (using appropriate categories and instances) or by repetition. As pointed out earlier, familiarity also may have a strong effect on feature elicitation. The more familiar an object is, the more features may be produced.[7]

It appears, then, that expressions which combine intermediate categories with higher-order categories ('the monkey is an animal', 'the monkey is a fish') will elicit more features for the *A-* than for the *B-term*. These intermediate categories can be an instance from either an appropriate or inappropriate higher-order category. Furthermore, expressions that combine higher-order categories with instances from these categories ('emotions are love') or instances from other categories ('emotions are roses'), will elicit fewer features for the *A-* than for the *B-term*.

Thus, the **fifth constraint on the stimulus selection is that expression types should not differ in the instance-category distribution over *A-* and *B-term***. However, the results of the expression categorization experiment (Section 4.4) show that subjects reliably differentiated among the expression types, only if a particular instance-category distribution coincided with an expression type. Therefore, instance-category distribution may be a decisive aspect for distinguishing among literals, metaphors, and anomalies.

Constraint 6: Equal syntax and fixed A- and B-term distribution

It is not self-evident that the compounds of an expression are analyzed in one unique way. The sentence 'the evening is a rooster...'[8] may be analyzed syntactically as subject - auxiliary verb - nomalized predicate (evening is rooster). However, the same sequence of words can also be read as nomalized predicate - auxiliary verb - subject, although the articles in the above example are strong cues for the first reading. Nonetheless, the second reading is possible, and in a metaphor with e.g., equal articles ('the raspberry is the heart', Table 4.0), the problem is less hypothetical.

In the second reading 'evening' becomes the nomalized predicate and is stressed by inversion (a well known poetic tool).[9] In that case, the metaphor may also be read as if a rooster is like the evening and not, as in the subject-first reading, the evening is like a rooster. In line with this idea, it is also not self-evident that the *A*- and *B-term* distribution follows the syntactical analysis, where subject = *A-term* and nomalized predicate = *B-term*. It could be argued that people sometimes read 'the evening is a rooster...' as subject - auxiliary verb - nomalized predicate, and still interpret the comparison between the *A*- and *B-term* the other way round. This would suggest that the *A*- and *B-term* distribution does not necessarily depend on grammar, and is influenced by other mechanisms as well. Table 4.0 gives three examples of ambiguous *A*- and *B-term* distributions, not reaching significance ($p > .05$) in a two-tailed sign test. How these data were obtained is described in Section 4.3.

It is also not necessarily, that *A*- and *B-term* distribution simply follows the plain order of appearance. Feature set size may be a factor of influence as well. Malgady & Johnson (1980) reported that 70% of the subjects preferred a low-salient term (poor of distinctive features) in first position, and a high-salient term (rich of distinctive features) in second position, as the right order of comparison for maximizing rated similarity in a metaphor. In other words, if a metaphor com-

Table 4.0: Three examples of ambiguous *A*- and *B-term* distribution (N = 10). Original versions in bold print.

	judged as *A-term*	
de stilte is jouw stem[10]	silence	30%
the silence is your voice	voice	70%
de framboos is het hart[11]	raspberry	70%
the raspberry is the heart	heart	30%
de regen is mijn bloed[12]	rain	70%
the rain is my blood	blood	30%

bines a high-salient term in first position, and a low-salient in second position, the order of comparison might just be reversed, despite the order of appearance or grammatical constraints. Although articles are expected to help determine *A*- and *B-term*, the number of activated features can interfere with this.

In other words, **the sixth constraint emphasizes that *A*- and *B-term* distribution ought to be fixed for all expressions (*A-term* first, *B-term* second)**, which can only be determined by testing. Section 4.3 discusses such a test.

Summary statements as to the effects of the basic linguistic properties are that orthography interacts with phonology, which sometimes interacts with meaning. Therefore, nonrhyming metaphors are selected from poems without systematic rhyme patterns (*constraint 1*). The number of syllables and letters influences readability, therefore, metaphors are selected that have monosyllabic *B-terms* with no more than four or five letters (*constraint 2*). Word frequency probably shows severe effects in *single term* conditions, so that metaphors are selected with *B-terms* in the same frequency domain (*constraint 3*). Lexical ambiguity may inhibit meaning selection, unless the ambiguity is dissolved by the context. However, in this stimulus set, lexical ambiguities will probably remain active, because poems are rather weak primers. Controls will be provided for deviant results (*constraint 4*). Category verification may be systematically correlated with the expression types, although this is not allowed (*constraint 5*). However, instance-category distribution may be a defining feature for different expression types. Finally, the selected metaphors should have the '*A* is *B*'-form, here, a nomalized predicate (two nouns). Since *A*- and *B-term* may not be distributed in agreement with syntactical analysis, their distribution has to be tested (*constraint 6*).

4.3 A priori stimulus selection on linguistic properties

Fifty one metaphors were selected from (fragments of) modern Dutch and Flemish poetry, written between 1920 and 1991. They consisted of two nouns, the second of which was monosyllabic and 4 or 5 letters long, in principle nonrhyming, without systematic spelling matches, and tentatively matched on word frequency.[13] The simple form '*A* is *B*' was chosen, so that the syntax of the expression would not suggest a metaphor. In a simile, the comparison is marked by the word 'like', which allows subjects to immediately perform a 'figurative strategy'. By omitting similes, subjects still may expect literal or anomalous endings. From the 51 metaphors, 51 literal expressions were constructed, by choosing a higher-order category for the second term, of which the first term was an instance (e.g., 'their house is a premises'). Fifty one anomalous expressions were formed, by substituting the *B-term* for a word that has meaning in one context only (e.g., in proverb like expressions and stock phrases), although it is not a nonword. Excluded from the selection were metaphors with rhyming or orthographically matching *B-terms*, such as 'de held is een speld' (the hero is a pin),[14] or 'de slaap is een slaaf' (the sleep is a slave).[15]

4.4 Test on the selection: Expression categorization and *A*- and *B-term* distribution

As described in the previous Section, 51 metaphors were matched on the relevant variables. For each metaphor, having the '*A* is *B*'-form, a literal and an anomalous version were constructed, by substituting the *B-term* for a word, which was supposed to make the expression literal or anomalous. This set of 153 expressions was doubled, by reversing the sentence order of each expression. Thus, all the expressions that had an '*A* is *B*'-form, also had a '*B* is *A*'-counterpart. The total list consisted of 306 expressions.

Five groups of subjects took part, all undergraduates in literature and modern languages, Dutch native speakers, between 20 and 28 years old. They volunteered in a selection task, in which they decided whether an expression was 'literal', 'metaphoric' or 'anomalous'. As a second task, subjects judged which of the two terms (*A* or *B*) was the focus in the expressions and which was the term of comparison.

The expressions were displayed on a monitor. Each expression was presented in two orders (never appearing after each other): 'the *A* is a *B*', and 'a *B* is the *A*'. This was done to avoid subjects from developing a simple strategy in the second task, namely, indicating the first term of all expressions as the *A-term*. All expressions were presented in pseudo-random order and stimulus sets were randomly distributed over subjects.

Subjects categorized the expressions by typing an 'L' for literal, 'M' for metaphor and 'A' for anomaly at cursor position, behind the expression. In the subsequent task, subjects underlined the focus term of the expressions, and put the term of comparison between brackets. In the instructions, expression types were exemplified by 'the monkey is an animal' for literals, and 'love is a rose' for metaphors. Anomalies were defined as neither literal nor metaphoric, but nonsensical expressions, without giving an example.

For the second task, the *A-term* was described as 'the focus term of an expression' and the *B-term* as 'the term that the focus term is compared with'. The words '*A-term*' and '*B-term*' were not used in the instructions. It was stressed that the focus term was not necessarily identical to the grammatical subject, to avoid that subjects merely performed a syntactical analysis.

After testing in a first group of 10 subjects, metaphors were removed that scored poorly ($p >> .05$) on category judgement or *A*- and *B-term* distribution in a two-tailed sign test. Their literal and anomalous counterparts were removed accordingly. Metaphors that were marginally significant ($p \approx .05$) were considered for retesting. Literal and anomalous expressions that did not reach significance were repaired. Metaphoric, literal and anomalous expressions were counterbalanced and the experiment was repeated with another group of 10 subjects as described above. This procedure was repeated exhaustively with more groups of 10 subjects, until three category sets were formed of 29 metaphors, 27 literals and 27 anomalies. In these last categories, two expressions fewer survived, because it was impossible to find words within the word frequency boundaries that were monosyllabic, not rhyming, and yet acceptable for the subjects. Fortunately, these missing expressions were the counterparts of metaphors with identical first terms ('death is a door' vs. 'death is a wall'; 'life is a bread' vs. 'life is a journey'). For the experiments described in Chapter 5 up to 7, this implies that these missing cases can be replaced by repeating a counterpart of a metaphor with an identical *A-term*.

Thus far, the conditions under which the expressions were tested, differed for every norm group. The number of expressions varied, because poorly performing expressions were removed, and the context for making the judgements changed, because other poorly performing expressions were altered. Finally, a group of 18 subjects judged the 29 metaphors, 27 literals and 27 anomalies that remained from the previous experiments. For the last two categories, two extra filler expressions were inserted, which were judged as literal or anomalous in the previous experiments, although the matching metaphor was not judged as metaphoric. In Table 4.1 the results for expression categorization, and *A*- and *B-term* distribution are presented for the last norm group of 18 subjects, showing that all selected expressions reached significance on both categorization and *A*- and *B-term* distribution. The Notes 16-44 show (fragments of) the poetic contexts.

Table 4.1: Expression categorization, and *A*- and *B-term* distribution results for *expressions*. Decisions for 'metaphor' are indicated by 'M', for 'literal' by 'L' and for 'anomaly' by 'A'. For all selected expressions p < .05, according to a two-tailed sign test (N = 18). For p < .001, rows are left empty. Superscripts refer to the literary sources.

Expressions	Translation	M	L	A	p =	focus	p =
de dood is een	death is a					death	
deur[16]	door	18				17	
feit	fact		17	1		18	
bubs	bunch	2		16	.002	17	
de dood is	death is					death	
een muur[17]	a wall	18				18	
het eind	the end	2	16		.002	18	
een kien	a bingo	1		17		17	
de dood is een	death is a					death	
reis[18]	journey	18				18	
ons hart is	our heart is					heart	
ons park[19]	our park	14		4	.030	17	
een spier	a muscle		18			18	
een zier	a whit	2		16	.002	17	
de haven is een	the harbor is a					harbor	
mond[20]	mouth	16	2		.002	18	
oord	place	1	17			18	
snars	know nothing	2		16	.002	17	
de haven is een	the harbor is a					harbor	
tuin[21]	garden	18				17	
plek	place	1	17			18	
plop	pop	1		17		17	
ons hoofd is een	our head is a					head	
berg[22]	mountain	16		2	.002	18	
ding	thing	1	17			17	
gooi	throw	3		15	.008	17	
mijn hoofd is een	my head is a					head	
haan[23]	rooster	14		4	.030	18	
deel	part	2	16		.002	17	
trant	trend			18		18	
de huid is een	the skin is a					skin	
vrouw[24]	woman	17		1		16	.002
laag	layer	2	16		.002	18	
taks	tax	2		16	.002	16	.002
hun huis is een	their house is a					house	
graf[25]	grave	17	1			17	
pand	premises		18			18	
wrik	jerk	1		17		17	

het kind is een	the child is a					child	
maan[26]	moon	14		4	.030	17	
mens	human being		18			18	
slip	slip	3		15	.008	17	
zijn lach is een	his laughter is a					laughter	
lied[27]	song	18				17	
klank	sound	3	15		.008	18	
prak	left over	2		16	.002	17	
het leven is een	life is a					life	
brood[28]	bread	17	1			18	
feit	fact	2	16		.002	18	
lurf	collar	2		16	.002	16	.002
het leven is een	life is a					life	
reis[29]	journey	18				18	
de maan is een	the moon is a(n)					moon	
dier[30]	animal						
ding	thing	16		2	.002	18	
piel	fumbler		18			17	
		4		14	.030	17	
de mens is een	man is a					man	
steen[31]	stone	17		1		18	
soort	species		18			18	
snap	snap	1		17		17	
deze middag is een	this afternoon is a					afternoon	
mand[32]	basket	14		4	.030	18	
tijd	time		15	3	.008	18	
krats	cheap buy	2		16	.002	17	
je mond is een	your mouth is a					mouth	
huis[33]	house	16			.002	18	
plek	place	3	15		.008	18	
rats	hotchpotch	2		16	.002	17	
de nacht is een	the night is a					night	
vrouw[34]	woman	18				17	
tijd	time		18			18	
dunk	estimate	2		16	.002	17	
de poëzie is een	poetry is a					poetry	
beer[35]	bear	16		2	.002	18	
taal	language	3	15		.008	18	
brui	give up	1		17		18	
poëzie is een	poetry is a(n)					poetry	
beest[36]	beast	17		1		18	
kunst	art	1	17			18	
klak	clack	2		16	.002	17	

de poëzie is een	poetry is a					poetry	
mens[37]	human being	18				15	.008
tekst	text	1	17			17	
snars	know nothing	2		16	.002	16	.002
de stilte is een	the silence is a					silence	
brug[38]	bridge	17		1		18	
feit	fact	2	16		.002	16	.002
lurf	confusion	1		17		17	
mijn toekomst is een	my future is a					future	
hart[39]	heart	17		1		18	
feit	fact		15	3	.008	18	
kluts	confusion	3		15	.008	18	
onze vrede is een	our peace is a					peace	
huid[40]	skin	16		2	.002	18	
recht	right	2	16		.002	18	
pats	slap	1		17		17	
de woede is een	rage is a					rage	
slang[41]	snake	17		1		17	
drift	temper	3	15		.008	18	
zwik	bunge	3		15	.008	17	
de zanger is	the singer is					singer	
zijn lied[42]	his song	14	3	1	.030	18	
een mens	a human being		18			18	
zijn keer	his turn	2		16	.002	16	.002
de zee is een	the sea is a					sea	
buik[43]	belly	17		1		17	
plek	place		17	1		18	
gros	gross	2		16	.002	16	.002
de zon is een	the sun is a					sun	
druif[44]	grape	16		2	.002	17	
ster	star	2	16		.002	18	
drom	mob	3		15	.008	17	

In Section 4.2 it was argued that *A-* and *B-term* distribution (which term is the *A-term*, which is the *B-term*?) does not necessarily follow syntactical analysis (subject = *A-term*, nomalized predicate = *B-term*), or the plain order of appearance (first term = *A-term*, second term = *B-term*). Expressions with an ambiguous focus ($p > .05$) were sifted out. They appeared to be expressions with either equal articles ('the wind is the tree'),[45] or combinations of definite articles and possessive pronouns (cf. Table 4.0, Section 4.2). Table 4.1 shows that the selected expressions mostly combine definite articles or possessive pronouns for the *A-term* with indefinite articles for the *B-term*. In these cases, the deictic element of the definite articles and possessive pronouns, and the syntactic subject in first position, coincide with the *A-term* focus. Therefore, probably no asymmetries between syntactical analysis and *A*- and *B-term* distribution will occur, which otherwise, might have obstructed the understanding of the expressions.

There are two exceptions to the demand of novelty in the metaphor set: 'Death is a journey' and 'live is a journey'. Although these metaphors are not idiomatic, they are already quite conventional. There are also two exceptions to the demand of free verse. The contexts of 'the harbor is a mouth' and 'the harbor is a garden' show end rhyme and sonnet form. Yet, analysis of the data in Chapter 5 and 6 without these stimuli did not alter the effects.

Steen (personal communication) stated that no checks are provided on the effects of the context on the expressions (Note 16-44). There may be unwarranted differences in the way *A*- and *B-term* are primed. Thus, the way in which the one expression is interpreted, may differ from the other, and therefore, obfuscate the results.

True as this may be, in the next Chapters, effects are controlled by quasi-F to see whether they are systematic on the stimulus level, also in *context*. If the effects that are predicted by the theories are not robust, they may be unimportant. On the other hand, these effects may be so subtle that uniformity of the contexts is a must (yet impossible to establish in natural texts).

In this study, *context* should be seen as a diffuse primer, in which the processing of the stimuli may differ from the more clear priming in *expression*. In a way, *context* could be viewed as adding extra noise to *expression* to see whether the effects still hold. Yet, if there are differences, it cannot be told exactly **which** aspects of *context* are responsible. The precise effects of *context* require much more preliminary research. The only question that will be answered here is whether there is a general tendency in adding poetic context or not.

4.5 Discussion for psychology: Post hoc controls of the constraints 3, 4, and 5

Word frequencies per expression type

Table 4.2 shows the mean word frequencies, which are calculated over all unique terms per term type (*A-term*, literal, metaphoric, and anomalous *B-term*). The values are sampled from the 42,380,000 words in the CELEX lexical database (Baayen, Piepenbrock & Van Rijn 1993). They consider the frequencies (*frq*) available for wordforms (CELEX list *dfw*), and their logarithmic transformation values (*lgv*), the latter of which emphasizes that the difference between a frequency of one and two is psychologically greater than between 1000 and 1001 (Burnage 1990: 3-101). Since the expressions were selected from natural texts (not designed for experimental purposes), certain words appear more often within a term type. For this reason, the number (N) of unique terms differs.

Table 4.2: Mean word frequencies (*frq*) and logarithmic transformation values (*lgv*) calculated over all unique terms per term type. Frequencies and logarithmic values are drawn from the 42,380,000 words in the CELEX lexical database.

Term type	*frq* mean	*sd*	*lgv* mean	*sd*	N
A-terms	7824.32	7444.00	2.03	.54	22
literal *B-terms*	6354.21	9662.39	1.73	.65	19
metaphoric *B-terms*	5191.77	7060.17	1.76	.54	26
anomalous *B-terms*	768.76	3601.55	.21	.56	25

Could the four term types in Table 4.2 be different word frequency groups, which might interfere with the responses predicted by the metaphor models? Or can the word frequency differences among the four term types be ascribed to random variation? A Oneway-ANOVA (Scheffé and LSD) showed that certain word frequency differences among the means of the four term types were not coincidental, neither for *frq* ($F_{3,88} = 4.38$, $p = .006$), nor for *lgv* ($F_{3,88} = 51.00$, $p = .000$). According to Scheffé, the effect on *frq* was due to the difference between *A-terms* and the lower-frequent anomalous *B-terms*. Least squares differences (LSD) - which allow between-groups differences to appear very quickly - found that all other term types were higher-frequent than the anomalous *B-terms*. With regard to *lgv*, both Scheffé and LSD indicated that anomalous *B-terms* were low-frequent compared to *A-terms*, literal and metaphoric *B-terms*. Comparable results were obtained while using the data for lemmas (CELEX list *dfl*) and two subsets from Uit den Boogaart (1975), each consisting of 120,000 words: Literary novels and short stories, and total written language.

Thus, systematic differences in word frequency were mainly caused by the anomalous *B-terms*, which can be regarded a separate word frequency group. Hence, word frequency may have an effect on the differences in feature set size and shared set size (Table 5.11, Appendix Chapter 5), reaction times (Table 6.3, Appendix Chapter 6), and electrocortical effects between the anomalous *B-terms* in comparison with all other term types.

It may be asked why these low-frequent words were used to create anomalous *B-terms*. The reason is simply that subjects - students of literature - recognized an intended anomaly as a metaphor as soon as a higher-frequent word was employed for the *B-term*. In other words, it was unavoidable to rely on more extreme cases to obtain a stable anomaly set.

Lexical ambiguities per expression type

What is the criterion for lexical ambiguity? Lexical ambiguous terms may be those words with more entrances (irrespective of part of speech) in the dictionary. In the Dutch dictionary Van Dale (1984), lexical ambiguities (homonyms) are indicated by Roman numerals. Table 4.3 shows the thus defined ambiguous terms in the expression set of Table 4.1.

Table 4.3: Lexical ambiguous terms in the stimulus set, as indicated by Roman numerals for homonyms in Van Dale (1984). L = literal, M = metaphor, A = anomaly. Translations of terms are found in Table 4.1.

A-terms	***B-terms***		
	L	**M**	**A**
dood	eind	deur	kien
hart	spier	muur	gooi
maan	ding	reis	taks
	deel	park	prak
	laag	berg	lurf
	pand	maan	piel
	klank	dier	snap
	recht	beer	krats
	ster	beest	rats
	oord	hart	dunk
		slang	klak
			kluts
			pats
			zwik
			gros
			drom

What may the effect be of ambiguous terms for the experiments to come? In the feature elicitation experiment (Chapter 5), average feature set size for ambiguities may be larger, because more meaning layers are present. Lexical ambiguous terms probably increase sentence verification times as well. Subjects may check which meaning of the ambiguous term is appropriate to the expression, before making a decision. Thus, lexical ambiguity must be controlled for in explaining variances for terms within expression type. These controls are found in Table 5.12 (Appendix Chapter 5), and Table 6.5 (Appendix Chapter 6).

Instance-category distribution per expression type

Instances of higher-order categories may be categories themselves for instances of a lower order. 'Botany' may be a higher-order category for the instances 'plants' and 'flowers'. 'Flowers', on the other hand, may be considered an intermediate category for the instances 'tulips' and 'roses'. In turn, 'roses' may be lower-order categories for the complete variety of existing roses. Theoretically, the subdivision is infinite and arbitrary. However, empirically, subjects prefer to use intermediate categories over higher and lower-order categories (Rosch & Mervis 1975).

Regarding the instance-category distribution of the expression types in Table 4.1, all *literal expressions* consist of an instance as *A-term* and an appropriate category as *B-term*. The literal *B-terms* are the widest categories in the expression set, such as 'fact', 'sort', 'time', and 'place'. This may be demonstrated by making a simple inversion. In 'the harbor is a place', all harbors are places, but not all places are harbors; in 'the sun is a star', all suns are stars, but not all stars are suns; in 'our heart is a muscle', all our hearts are muscles, but not all muscles are hearts. The *A-terms* of the expressions thus can be regarded as intermediate categories, which are appropriate instances of the higher-order categories of the *B-terms*. Hence, processing the literal expressions will probably involve category verification from an instance to a category.

Processing the *metaphors* (e.g., 'my head is a rooster') will probably involve the matching of an instance in the *A-term* ('head') with an instance from an inappropriate category in the *B-term* ('rooster'). Since instances from higher-order categories ('body parts' and 'animals') are intermediate categories themselves, two inappropriate intermediate categories are matched.

The compounds of the *anomalies* are an intermediate category for the *A-term*, which must be compared with an inappropriate instance for the *B-term*. This instance is very specific, in that it has a strong contextually bound meaning in the Dutch language. It is hard to think of any more specific instances belonging to these lower-order categories, because the words themselves are almost meaningless.

In short, Table 4.1 shows literal expressions that match an instance with an appropriate higher-order category, metaphors that match two inappropriate instances (or two inappropriate intermediate categories), and anomalies that match an intermediate category with an instance from an inappropriate lower-order category.

For the feature elicitation experiments, discussed in Chapter 5, this probably means that the *A-terms* and metaphoric *B-terms* evoke the largest number of features. According to Rosch (1973), the intermediate categories are the best known, and are used most often. It could be argued, then, that subjects will sum up the most features for what they know best. Since the anomalous *B-terms* (lower-order categories) are the least frequently used words (cf. Table 4.2), they probably evoke the least features. The higher-order categories of the literal *B-tarms* supposedly take a position in between, evoking moderate numbers of features. Since the literal expressions are the only stimulus type with appropriate instances and categories, instance and category dominance will be highest for these expressions, probably affecting the response times (cf. Balota & Chumbley 1984) in the RT experiments (Chapter 6). The distributions of instance-category dominance are found in Table 5.13, Chapter 5.

4.6 Discussion for the theory of literature

The research discussed in Section 4.2 was not designed to investigate questions raised by the theory of literature. Nonetheless, the results can shed some light on certain issues. The experiments discussed in Section 4.4 sorted out which expressions were considered metaphors, literals or anomalies, and how *A*- and *B-term* were distributed. In this Section, the ins and outs of the selected expressions presented in Table 4.1 will be discussed, and aspects which are of relevance to the theory of literature will be stressed.

Reconsidering spelling and rhyme

Word pairs, similar to those suggested by Wellek & Warren (1955: 161-162) (heart: part, tears: fears), were investigated by Polich, McCarthy, Wang & Donchin (1983), and Kramer & Donchin (1987). These researchers studied word pairs such as 'match-patch', 'blare-stair', 'catch-watch', 'shirt-witch', which differed systematically in the congruency between the rhyme- and spelling-dimension, thereby, emerging into 'the special problem of the eye-rhyme' (Wellek & Warren 1955: 162).

The results of the studies by Polich et al. and Kramer et al. indicated that the recognition of rhyme in such word pairs depended foremost on the recognition of spelling matches (the eye-rhyme), after which the phonological rhyme was recognized. Thus, it may be concluded that reading rhyme in writing is probably not the matching of words in a pure phonological comparison, but also an implicit spelling or typographical check in an orthographical comparison.

On the question whether rhyme interacts with meaning, studies by Sanquist, Rohrbaugh, Syndulko & Lindsley (1980), Rugg (1984), Rugg & Barrett (1987), and Hoorn (1996) might be of interest, in showing that particularly rhyme mismatches (nonrhymes where rhymes were expected) interacted with semantics. This is interesting, because from the point of view of Wellek & Warren (1955) and Jakobson (1981), the largest interactions with semantics are to be expected in rhyming conditions, not in nonrhyming conditions.

Reconsidering meter as syllable count

Jakobson (1961: 252) reported that already in the beginning of this century, statistics were employed for the investigation of poetry. This concerned the count of syllables as a predictor for anticipation and surprise.

> The Russian school of metrics owes some of its internationally echoed achievements to the fact that some forty years ago such students as B. Tomashevskij, expert both in mathematics and philology, skillfully used Markov chains for the statistical investigation of verse; these data, supplemented by a linguistic analysis of the verse structure, gave in the early twenties a theory of verse based on the calculus of its conditional probabilities and of the tensions between anticipation and unexpectedness as the measurable rhythmical values, and the computation of these tensions, which we have labeled 'frustrated expectations', gave surprising clues for descriptive, historical, comparative, and general metrics on a scientific basis (Jakobson 1961: 252).

In Wellek & Warren (1955), a review is given of these statistical methods. However, their criticism of these methods is focused on the neglect of the semantic aspect of detecting meter. Wellek & Warren (1955) state:

> We may be doubtful about a good many features of the Russian theories, but one cannot deny that they have found a way out of the impasse of the laboratory on the one hand, and the mere subjectivism of the musical metrics on the other. Much is still obscure and controversial; but metrics has today restored the necessary contact with linguistics and with literary semantics. Sound and meter, we see, must be studied as elements of the totality of a work of art, not in isolation from meaning (Wellek & Warren 1955: 176).

How then, does this Russian meter statistics work, and how can it be connected to a 'meaning count'? To start with the first question, Wellek & Warren (1955: 174) stated that 'the statistical method used is very simple'.

> In each poem or Section of a poem to be analyzed, one counts the percentage of cases in which each syllable carries a stress. If, in a pentameter line, the verse should be absolutely regular, the statistics would show zero percentage on the first syllable, 100 per cent on the second, zero on the third, 100 on the fourth, etc. This could be shown graphically by drawing one line for the number of syllables and another, vertically opposed to it, for the percentages. Verse of such regularity, is of course, infrequent, for the simple reason that it is extremely monotonuous. Most verse shows a counterpoint between pattern and actual fulfillment e.g., in blank verse the number of cases of accents on the first syllable may be rather high, a well-known phenomenon described either as the "trochaic foot," or "hovering" accent, or "substitution." In a diagram, the graph may appear flattened out very considerably; but if it is still pentameter and intended as such, the graph will preserve some general tendency toward culmination points on syllables 2, 4, 6, and 8. This statistical method is, of course, no end in itself. But it has the advantage of taking account of the whole poem and thus revealing tendencies which may not be clearly marked in a few lines. It has the further advantage of exhibiting in a glance the differences between schools of poetry and authors (Wellek & Warren 1955: 174).

The Russian meter statistics, in short, can be expressed as the average number of similar stress patterns (the 'feet') per sentence of a poem. This 'meter value' of the poem, obtained by the syllable count of the Russian Formalists, may be combined with a syllable count that expresses certain semantical aspects of the poem, for instance, its readability, or understanding ease. Flesh (1948) developed a measure for readability, which expresses the reading ease of a text as a measure of the number of syllables per 100 words and the number of words per sentence. The validity of Flesh's formula for real readers is rather high, as investigated by Stevens & Stone (1947). Flesh pointed out that word length is an indirect measurement of word complexity, which is in turn an indirect measurement of abstraction. Generally speaking, the more syllables the word has, the more the abstraction level of the word is rated as complex by the reader.

How can these two syllable-counts - the Russian meter statistics and Flesh's readability measure - be combined in a single study? Suppose, a connection is assumed between a formal aspect of poetry e.g., a certain verse foot (the anapaest '◡◡–' versus the dactyl '–◡◡'), and a semantic aspect, its abstraction level. The research could start with a statistical inventory of the stress-patterns in poetry, which should result in a group of poems with a high average of anapaest meter for every poem, and a group of poems with a high average of dactyl meter for every poem. A second statistical inventory describes the average readability for every poem, as described by the Flesh (1948) formula. Correlations could now be calculated between calculated readability and calculated meter type, to see which meter type is connected to what abstraction level.

An empirical control for this statistical correlation can be provided by two groups of readers, one of which rates the readability of both the anapaest and dactyl poems, and the other reads aloud these poems, while an oscilloscope registers the stress patterns of the voice. The statistic inventory of verse feet should now highly correlate with the voice stress patterning. Similarly, the calculated readability should be highly correlated with the reader's rated readability. Moreover, in a regression analysis, the correlation between calculated readability and calculated meter type should be similar to the correlation between rated readability and voice stress patterning, to supply a full control of the assumed connection between meter and abstraction level in semantics.

Klare (1963) provided a review of readability formulas, among which Flesh's is one of the most popular. In connection to the study of literature, Note 3 at the bottom of this Chapter could be of interest, where a sample of literary metaphors is presented with a systematic difference in number of letters and syllables of the *B-term*. An extended and adapted version of this set could be used to further investigate the correlation between number of syllables and degree of complexity, as predicted by Flesh's formula.

A connectionist model of meter is proposed by Hayward (1996), in which the statistical (ir)regularities of poetic meter are analyzed by the computer. With a fair degree of success, this model is capable of distinguishing, for instance, Pope and Tennyson from Donne and Wordsworth in the way they use poetic meter.

Reconsidering word frequency

Word frequency - the number of times a word is used by a speech community - is of interest to the comparison and anomaly view on metaphor processing. These theories claim that the *C-term* of a metaphor can be predicted from the single words. Seldomly used words are more difficult to recognize than highly familiar words. The expressions in Table 4.1 are checked and matched as much as possible on word frequency, because if the literals would use very low-frequent words and the metaphors only high-frequent ones, the differences in processing between the two expression types could be explained in terms of word frequency, and not in terms of the special metaphoric processes that the theories assume.

Furthermore, things of which we know little probably elicit fewer associations (features) than familiar things. Thus, it could be argued that word frequency is correlated with feature elicitation, because frequency of use might indicate a degree of familiarity to the speech community. Again, differences in word frequency could lead to differences in the number of features. In turn, this would have consequences for the metaphor theories in that the chance to form a *C-term* changes. If high frequency of use is connected with strong feature activation, the features of an *A-term* have higher probabilities to be matched with *B-term* features, because the *B-term* is feature-rich. Furthermore, a metaphor with a high-frequent *B-term* would be processed easier than a metaphor with a low-frequent *B-term*. Things could get complicated, if the processing of an anomaly with a high-frequent *B-term* would be compared with the processing of metaphors with low-frequent *B-terms*.

Psychological research discussed in Section 4.2 showed that, indeed, word frequency has an effect on word recognition. Statistical analysis in Section 4.5 on the word frequencies of the term types in Table 4.1 showed that only the anomalous *B-terms* can be regarded as a separate word frequency group, which may influence the number of elicited features (Table 5.11, Appendix Chapter 5), reaction times (Table 6.3, Appendix Chapter 6) and EEG.

Reconsidering lexical ambiguity (homonymy)

Wellek & Warren (1955: 162) stated that 'the rhyming of homonyms (...) is a form of punning'. Lexical ambiguities are words with multiple meanings, and actually almost any word is somewhat ambiguous. Usually, context primes the appropriate meaning, but in literature, this priming is often left open, thus letting multiple meanings remain active. The psychological literature discussed in Section 4.2 showed that unscrambling an unprimed ambiguity influences the processing of a word.

This effect of lexical ambiguity will probably be found in every condition in the experiments to come. Table 4.3 in Section 4.5 summarizes the lexical ambiguities in Table 4.1, showing that the anomalous *B-terms* count the highest number of lexical ambiguities. As to the effects of ambiguity in understanding these expressions, ad hoc controls will be performed in Table 5.12 (Appendix Chapter 5) and Table 6.5 (Appendix Chapter 6).

Another point of interest prompted by the psychological research into lexical ambiguity is that in the early encoding of a word, all meanings of the word become active - independent of contextual bias - after which the appropriate meaning is selected with the help of contextual cues. The comparison and anomaly view assume that the encoding of a word results into a fixed feature set. All meanings are activated, independent of context. The interaction view, on the other hand, emphasizes the role of context in priming the meaning of a word. Since none of the theories describe the time series of the encoding phase (or any other phase for that matter), it is argued in

Section 4.2 that both points of view could be right. The comparison and anomaly view may describe the initial, early, encoding of words, in which the fixed feature sets are automatically activated. This would support the ideas of Šklovskij (1965) that normal speech activates highly automatized meanings. The interaction view would describe the end product of this encoding, where a rapid selection from the automatically accessed meanings takes place. This selection would be sensitive to contextual information, world and genre knowledge, which is in line with Šklovskij's idea that the automatized meanings are renewed and made unfamiliar in literature.

Reconsidering categories and groups of objects

Wellek & Warren hinged the category membership of words as a part of their semantic nesting; 'whether, for example, they belong to one or several linguistic categories (parts of speech, different cases) or groups of objects' (Wellek & Warren 1955: 161-162). In other words, when a word is a category name, then the 'semantic sphere' of that word, consisting of a feature set, will probably contain an appropriate instance. Reversely, when the word is an instance, the 'semantic sphere' (the feature set) will probably contain the category name. The probability that an instance is associated with a category name is called instance dominance. The probability that a category name is associated, given the instance, is called category dominance. Instance and category dominance can be either high or low, and is expressed by the average naming frequency of instance or category in a feature elicitation experiment (Table 5.13, Chapter 5).

Psychological research as discussed in Section 4.2 showed that high category and instance dominance in word combinations led to quick processing of these combinations. The expressions shown in Table 4.1, are also word combinations in which category verification (are instance and/or category appropriate?) might play a role.

Psychological research also made clear that there are three types of categories that people use to organize the world. Lower-order categories (e.g., 'panegyrics'), which are instances of intermediate categories (e.g., 'poetry'), which are instances of higher-order categories (e.g., 'literature'). The intermediate categories are most frequently used by the speech community. In Section 4.2, it is argued that they can be assumed to elicit the highest feature rates, and that they are processed the fastest. It is necessary, therefore, to analyze the expression set of Table 4.1 (Section 4.5), to find out if an expression type is systematically connected to a specific instance-category distribution.

Table 4.1 shows literal expressions that match an instance for the *A-term* with an appropriate higher-order category for the *B-term* (e.g., 'this afternoon is a time', 'rage is a temper'). The metaphors consist of two inappropriate instances, or, put differently, two inappropriate intermediate categories. The anomalies match an intermediate category for the *A-term*, with an instance from an inappropriate lower-order category for the *B-term*, which is a very specific word, extracted from a proverb or stock phrase.

As discussed in Section 4.5, this probably results in high feature rates for the literary metaphors, which use the best known, intermediate categories. Intermediate feature rates will probably be found for the literal expressions, which use less well known, higher-order categories for the *B-term*. The least features will be activated by the anomalies, which combine a very specific instance of even less well known, lower-order categories (here, usually stock phrases) as a *B-term*.

Since the literal expressions are the only stimulus type that combines appropriate instances and categories, instance and category dominance will be the highest for this expression type, which also might have an effect on the response times (see Balota & Chumbley 1984) in the RT experiments discussed in Chapters 6 and 7.

Notes:

1. A Homeric comparison combines several metaphors on different levels, often embedded in the main sentence:

> Like the sheep of the heath, pass
> Late through the green evening light, so that who stands
> On a moss-grown hill, sees them go
> From the moor-edge and into a dark alley,
> Around the corner - thus left that dark troop
> Which she stood watching as long as the cries
> Of birds still rose, the outstretched beach.

The above (own) translation is from the Dutch:

> Zooals de schapen van de heide, laat
> Door 't groene avondlicht gaan, dat wie staat
> Op een bemosten heuvel, ze ziet gaan
> Van den heizoom en in een donkre laan,
> Den hoek om - zoo verliet die donkre troep
> Die zij nazag zoolang nog het geroep
> Van vogels opging, het gerekte strand.

Gorter, H. (1940). Mei, een gedicht. In: *Herman Gorter, Mei, een gedicht*. P.N. van Eyck (Ed.). Elsevier, Amsterdam, 21.

2. 'Een vlies van mist'. Nijhoff, M. (1993). Midzomer. In: *Gedichten, Deel 1, Teksten*. W.J. van den Akker & G.J. Dorleijn (Eds.). Van Gorcum, Assen, Maastricht, 228.

3. In the table of metaphor types drafted by Genette (1972: 30), this kind of metaphor is classified as 'not motivated identifications'. These consist of *A*- and *B-term*, without 'like' or 'as'. Unlike the examples below, they also lack adjectives or verbs that might motivate the metaphor:

'My eye is an insatiable spider' (mijn oog is een onverzadigbare spin). Lucebert (1980). Parcival. *Val voor vliegengod*. De Bezige Bij, Amsterdam, 35.

'The eye is scientifically a repulsive apparatus' (Het oog is wetenschappelijk een weerzinwekkend apparaat). Lucebert (1993). Het oog. *Van de roerloze woelgeest*. De Bezige Bij, Amsterdam, 11.

4. Schoonheid als wapen tegen oorlog is een zwaard van ijs. Enquist, A. (1991). Chambord. *Soldatenliederen*. Amsterdam, De Arbeiderspers, 25.

5. This finding could be tested for the comprehensability of *B-terms*, with the set of literary metaphors presented below. These metaphors are ordered from the *B-term*'s smallest number of syllables to the *B-term*'s largest number of syllables. Within an equal-syllable group, metaphors are ranked from the smallest number of letters to the largest. The *B-terms* in the top section should thus be judged as easier to comprehend than the ones in the bottom section of the list. For a more sophisticated test, equal replications for each group should be found, matched on the relevant variables as discussed in Section 4.2.

One syllable, three letters width:

God is een pop, de duivel een mot	(God is a puppet, the devil a moth)
mijn keel is een pad	(my throat is a trail)
de vaalt is zijn zee	(the dunghill is his sea)
de dichter is een koe	(the poet is a cow)

One syllable, four letters width:

de dag is zand	(the day is sand)
de hond is taal	(the dog is language)
de wind is de boom	(the wind is the tree)
het huis is een veld	(the house is a field)
angst is de veer	(fear is the spring)
de smart is een ring	(grief is a ring)
de etna is de zuil	(the etna is the column)
de nevel is een boom	(the haze is a tree)
de herder is de deur	(the shepherd is the door)
onze vreugde is een huis	(our joy is a house)
uw gezicht is een haan	(your face is a rooster)
het gedicht is een zwam	(the poem is a fungus)
poëzie is een cent	(poetry is a penny)
de ontzetting is de gast	(the appal is the guest)
de ontzetting is een stad	(the appal is a town)
de ontzetting is een kies	(the appal is a molar)

One syllable, five letters width:

hun oog is een steen	(their eye is a stone)
mijn mond is nacht	(my mouth is night)
jouw ziel is beeld	(your soul is image)
de angst is een lucht	(the fear is a sky)
de Lente is 'n feest	(spring is a feast)
driekantige dame is een zwijn	(three-sided lady is a swine)
de ontzetting is de bruid	(the appal is the bride)

Two syllables, four letters width:

mijn stap is de adem	(my step is the breath)
de gedachte is mijn echo	(the thought is my echo)

Stimulus selection:

Two syllables, five letters width:

de dag is een wagen	(the day is a coach)
de mens is mijn koude	(man is my cold)
een hand is een leven	(a hand is a life)
het hart is een hamer	(the heart is a hammer)
je stem is een aster	(your voice is an aster)

Two syllables, six letters width:

de zee is moeder	(the sea is mother)
de koude is mijn kennis	(my cold is my knowledge)
de liefde is een moeder	(love is a mother)
het licht is de liefde	(the light is the love)
zijn wekker is oorlog	(his alarm-clock is war)
Eenzaamheid is een woning	(lonelyness is a dwelling)

Two syllables, seven letters width:

de maan is de spiegel	(the moon is the mirror)
poëzie is een stuiver	(poetry is a penny)
de ontzetting is een bloedbad	(the appal is a massacre)
de stem is een kiesschijf	(the voice is a dial)

Three or more syllables, seven or more letters width:

onze duisternis is een obelisk	(our darkness is an obelisk)
het water is een winterdas	(the water is a winter tie)
het badzand is nu kolengruis	(the bath sand is now coal-dust)
de zon is een hatende catalaan	(the sun is a hating catalonian)
de duivel is een fonofaag,	(the devil is a phonophage,
een cacafonofaag	a cacaphonophage)

Literary sources of the above metaphors (in order of appearance):

Claus, H. (1966). In het museum van Chicago. *Gedichten 1948-1963*. De Bezige Bij, Contact, Amsterdam, Antwerpen, 216.

Lucebert (1974). Nympholalie. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 235.

Ter Balkt, H.H. (1982). De vaalt. *Machines! Maai ons niet, maai de rogge. Bloemlezing gedichten 1969-1979*. De Harmonie, Amsterdam, 12.

Achterberg, G. (1985). De dichter is een koe. *Verzamelde gedichten*. Querido, Amsterdam, 101.

Campert, R. (1976). Vogels vliegen toch. *Alle bundels gedichten*. De Bezige Bij, Amsterdam, 33.

Elburg, J.G. (1975). Geen tijd. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 237.

Schierbeek, B. (1978). De derde persoon. *Het boek ik; De andere namen; De derde persoon*. De Bezige Bij, Amsterdam, 438.

Ter Balkt, H.H. (1978). 8 Het huis. *Waar de burchten stonden en de snoek zwom*. De Harmonie, Amsterdam, 73.

Lodeizen, H. (1969). Angst. *Nagelaten werk, gedichten 1948*. Van Oorschot, Amsterdam, 36.

Lodeizen, H. (1969). De hemel is een cirkel rondom mijn ellende. *Nagelaten werk, gedichten 1948*. Van Oorschot, Amsterdam, 22-23.

Schierbeek, B. (1981). De gestalte der stem. *De gestalte der stem; Het dier heeft een mens getekend; Ezel mijn bewoner*. De Bezige Bij, Amsterdam, 40.

Campert, R. (1976). Een dag op aarde. *Alle bundels gedichten*. De Bezige Bij, Am-

sterdam, 71.
Schierbeek, B. (1981). Het dier heeft een mens getekend. *De gestalte der stem; Het dier heeft een mens getekend; Ezel mijn bewoner*. De Bezige Bij, Amsterdam, 155.
Lucebert (1974). V. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 125.
Lucebert (1974). Romance. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 33.
Claus, H. (1966). Spreken 10. *Gedichten 1948-1963*. De Bezige Bij, Contact, Amsterdam, Antwerpen, 185.
Habakuk II de balker (1970). Poëzie. *Uier van het oosten*. De Bezige Bij, Amsterdam, 43.
Ter Balkt, H.H. (1986). De ontzetting. *Verkeerde Raadhuizen*. Nova Zembla, Arnhem, 34.
Ter Balkt, H.H. (ibid.).
Ter Balkt, H.H. (ibid.).
Nooteboom, C. (1984). Een doorregende landstreek. *Vuurtijd, ijstijd, gedichten 1955-1983*. De Arbeiderspers, Amsterdam, 159.
Schierbeek, B. (1978). De derde persoon. *Het boek ik; De andere namen; De derde persoon*. De Bezige Bij, Amsterdam, 402.
Simpelaar, P. (1975). De Elyzéese velden van Zeeland. *Verzamel de wolken op je gemak*. De Arbeiderspers, Amsterdam, 54.
Andreus, H. (1975). De sonnetten van de kleine waanzin, 3. *Gedichten 1948-1974*. Uitgeversmij. Holland, Haarlem, 112.
Van Ostaijen, P. (1952). Music-Hall. *Verzameld werk / poëzie*. De Sikkel, Daamen, Van Oorschot, Amsterdam, 13.
Simpelaar, P. (1975). Wees gerust 't is zoet volk. *Verzamel de wolken op je gemak*. De Arbeiderspers, Amsterdam, 21.
Ter Balkt, H.H. (1986). De ontzetting. *Verkeerde Raadhuizen*. Nova Zembla, Arnhem, 34.
Van Ostaijen, P. (1952). Herinnering. *Verzameld werk / poëzie*. De Sikkel, Daamen Van Oorschot, Amsterdam, 64.
Elburg, J.G. (1975). Astarte 1. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 237.
Elburg, J.G. (1975). Détour de france. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 125.
Elburg, J.G. (1975). Leven is in strofen geschreven 2. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 198.
Marsman, H. (1979). De hand van de dichter. *Verzamelde gedichten*. Querido, Amsterdam, 69.
Rodenko, P. (1975). Besneeuwd landschap, III, Woorden van brood. *Tulpensnijder, orensnijder, verzamelde gedichten*. De Harmonie, Amsterdam, 81.
Lucebert (1974). V. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 125.
Hillenius, D. (1971). De zee is een buik vol embryos. *Tegen het vegetarisme*. Van Oorschot, Amsterdam, 46.
Elburg, J.G. (1975). Leven is in strofen geschreven 2. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 198.
Elburg, J.G. (1975). Om met penselen op te tekenen. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 191.
Lucebert (1974). Tajiri. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 204.
Elburg, J.G. (1975). Leven is in strofen geschreven 2. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 198.
Claus, H. (1966). Het klemwoord: huis, 12. *Gedichten 1948-1963*. De Bezige Bij, Contact, Amsterdam, Antwerpen, 164.
Elburg, J.G. (1975). Om met penselen op te tekenen. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 191.
Habakuk II de balker (1970). Poëzie. *Uier van het oosten*. De Bezige Bij, Amsterdam, 43.

Ter Balkt, H.H. (1986). De ontzetting. *Verkeerde Raadhuizen*. Nova Zembla, Arnhem, 34.
Vinkenoog, S. (1972). Een wereld: audio-visueel gedicht. *Wonder boven wonder*. De Bezige Bij, Amsterdam, 29.
Lucebert (1974). V. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 125.
Lucebert (1974). Het orakel van monte carlo. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 112.
Lucebert (ibid.).
Elburg, J.G. (1975). Ik zeg: de zon is een druif, nee een meloen. *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 126.
Lucebert (1980). Voorwoord voor val voor vliegengod. *Val voor vliegengod*. De Bezige Bij, Amsterdam, 11.

6. Word frequency may not affect category verification times, but it could still be influential on surprise effects (cf. Section 4.1 on word frequency).

7. Conclusively, it could be argued that high frequency of a word is positively correlated with high familiarity values and strong feature elicitation, but negatively with number of syllables.

8. 'De avond is een haan, vanwege driemaal Petrus'. Claus, H. (1966). Tancredo Infrasonic 8. *Gedichten 1948-1963*. De Bezige Bij, Contact, Amsterdam, Antwerpen, 107.

9. For instance:

'A black discotheque is the sleep' (een zwarte discotheek is de slaap). Vinkenoog, S. (1966). Tweespraak. *Eerste gedichten 1949-1964*. De Bezige Bij, Amsterdam, 167.

'A sigh is man' (een zucht is de mens). Vinkenoog, S. (1972). Drugsong. *Wonder boven wonder*. De Bezige Bij, Amsterdam, 97.

'A wandering bird is the poet' (Een zwervende vogel is de dichter). Vinkenoog, S. (1972). Strijdlied en liefdesgedicht. *Wonder boven wonder*. De Bezige Bij, Amsterdam, 136.

'O a noise filter is death' (o een ruisfilter is de dood). Schierbeek, B. (1981). Het dier heeft een mens getekend. *De gestalte der stem; Het dier heeft een mens getekend; Ezel mijn bewoner*. De Bezige Bij, Amsterdam, 45.

10. Habakuk II de balker, (1987). Panter en krekel (Script). *Aardes deuren*. De Harmonie, Amsterdam, 15-16.

11. Schierbeek, B. (1978). De derde persoon. *Het boek ik; De andere namen; De derde persoon*. De Bezige Bij, Amsterdam, 518.

12. Claus, H. (1970). 17. Heer everzwijn, zijn nota's bij "Genesis I, 1". *Gedichten*. De Bezige Bij, Amsterdam, 61.

13. A more extended analysis of the word frequency distribution in the stimulus set is provided by Table 4.2, after the categorization of the expressions by the subjects.

14. Kouwenaar, G. (1968). Zoals Van Speyck zei:. *De stem op de derde etage*. Querido, Amsterdam, 50.

15. Vinkenoog, S. (1972). *Wonder boven wonder, gedichten 1965-1971*. De Bezige Bij, Amsterdam.

16. Vinkenoog, S. (1972). Drugsong. *Wonder boven wonder, gedichten 1965-1971*. De Bezige Bij, Amsterdam, 100.

Drugsong 2

van wat ons hier vandaag tesamenbrengt:
een zingende muze, muziek die ontspringt
aan liefde en kennen en herkennen,
ach, ware het leven zo zoet
als de dood die ons allen verenigt -

Niets kon ik haar zeggen,
niets dat haar kon troosten
of op weg kon helpen -
want de dood is een reis
die ieder zelf moet maken,
de dood is een deur
die je zélf moet ontsluiten...

17. Andreus, H. (1975). Muur. *Gedichten 1948-1974*. Uitgeversmij. Holland, Haarlem, 79-80.

Muur

Wij hebben een holte in onze borst
van gemis, wij en eenieder, maar
wie waagt zich daaraan?
Wie leeft zich ten dode uit?
Alleen de lafste leugens
worden rijkelijk betaald.

Wat moeten wij met leven
voor die muur en het vraag-
teken erna?

Liefde werd ons genomen
zoals men een appel schilt,
Uit. De dood is een muur.

18. Vinkenoog, S. (1972). Drugsong. *Wonder boven wonder, gedichten 1965-1971*. De Bezige Bij, Amsterdam, 100.

Drugsong 2

van wat ons hier vandaag tesamenbrengt:
een zingende muze, muziek die ontspringt
aan liefde en kennen en herkennen,
ach, ware het leven zo zoet
als de dood die ons allen verenigt -

Niets kon ik haar zeggen,
niets dat haar kon troosten
of op weg kon helpen -
want de dood is een reis

19. Campert, R. (1976). Zwanen. *Alle bundels gedichten*. De Bezige Bij, Amsterdam, 57.

Zwanen

Er wordt geschreven aan vele boeken;
in groepjes wandelen we pratend naar een nieuw hoofdstuk toe
of ook wel alleen. En hoe makkelijk niet
springen we van de ene alinea op de andere.

Ons hart is ons park, nietwaar?
en ons lichaam een mooi jong traliehek
en zie: zwanen, schitterend van vuil, drijven
in onze stinkend geprezen maag.

20. Snoek, P. (1982). De haven, II. *Verzamelde gedichten*. Manteau, Antwerpen, 655.

De haven is een mond
die drinkt aan alle glazen,
waar vreemde schepen grazen
als koeien op een fond

van groen en in het rond,
op stalen boegen, lazen
wij land, zonder verbazen,
dat nooit zijn grenzen vond.

Er liggen vastelanden
en bloed van vele handen
gestapeld op de ree,

die als een kleine aarde
het onbegrip bewaarde
dat aanwaaide uit de zee.

21. ibid. III, 656.

De haven is een tuin,
waar alle mooie kleuren
en alle sterke geuren
gedijen in een puin

vol vreemde rassen. Bruin
zijn openstaande deuren,
waar hoerenwaarden leuren
met vrouwen zacht als schuim.

Een kind geeft witte rozen
met vreugde in, aan matrozen,
die dronken zijn en rood.

Maar allen daar vergeven
de vreemdeling, wiens leven
zal reven naar de dood.

22. Schierbeek, B. (1981). Ezel mijn bewoner. *De gestalte der stem; Het dier heeft een mens getekend; Ezel mijn bewoner*. De Bezige Bij, Amsterdam, 289.

onze handen zijn zo klein dat zij niet kunnen vallen
niemand zou onze handen horen vallen
zo klein zijn onze handen
dat is over 100.000 jaar
maar ons hoofd is een berg vol geweld
gestileerd naar het ei
en net zo ingewikkeld
de ongeziene draden die er lopen zien wij
de relaties en het schaduwpatroon van ons krachtveld
onzichtbaar

23. Lucebert (1974). gedicht voor een zeer hoofd. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 327.

gedicht voor een zeer hoofd

de wind verft mijn ogen om
tot spitse witte vlaggen
en ik geef mijn hoofd over
aan de grote verre wolken

in de grote verre wolken
zitten maanzieke honden
als door de ramen de zon schijnt
zijn de honden zonzieke poezen

mijn benen jengelen uit de verre wolken
als van processies verstoken klokken
maar mijn hoofd is een haan tussen de honden
en brult en balkt en blaast en blaft
terwijl de honden huilen

24. Schierbeek, B. (1978). De derde persoon. *Het boek ik; De andere namen; De derde persoon*. De Bezige Bij, Amsterdam, 440.

o heer ik schilder huizen en ik wil van het hart van de aarde
van u los zijn o heer
en van het hart van de aarde de riem snijden
die mij mijn middel meet o heer
want weet ik bid met een dier om de aanwezigheid
in ben o heer als de huid van een rubberboom zo menselijk
en de huid is een vrouw
en de vrouw ging uit
zij staat op de gloeiende brug van het zijn
met de kapotte stad die zij is
in de geroosterde menselijke huid die ik zing

25. Nooteboom, C. (1984). Een doorregende landstreek. *Vuurtijd, ijstijd. Gedichten 1955-1983*. De Arbeiderspers, Amsterdam, 159.

Een doorregende landstreek

Een doorregende landstreek
een verroest zwaard. Een opgebruikte mythe.
De onsterfelijken zijn gestorven en vergeten,
hun huis is een graf.
Hun oog is een steen waar ze alles mee zien:

dit punt en zijn afstand,
en alle lengtes van tijd daartussen -
de wellust die hun lichaam steeds verder insluit
in een corrupte beweging.

26. Schierbeek, B. (1981). Ezel mijn bewoner. *De gestalte der stem; Het dier heeft een mens getekend; Ezel mijn bewoner*. De Bezige Bij, Amsterdam, 37.

maar ook en hetzelfde is:
  dat de vogel zingt
  de tak breekt
  de vogel valt
  en de boom ontwaakt
ik zeg
  als de ochtend valt
  op het eerste gezicht van de zee
  ziet het strand groen
  en zingen wit de dode bloemen
want
  het kind is een maan
  een sikkel in morgen en nacht
  en verdeelt de stam
  tot kruin van verwarring
  maar het kind spreekt
  en praat verder

27. Marsman, H. (1975). Seine-et-Marne. *Verzamelde gedichten*. Querido, Amsterdam, 60.

Tussen Marlotte, Montigny en Grez
staat op een heuvel een klein houten huis.
een stem vraagt loom: is de zon wit of grijs?
wordt de dag snikheet?
en de stem van den jongen die buiten staat
en de reis verspiedt van het vroege licht,
lacht en zijn lach is een lied.

28. Coninck, H. de (1986). Met de vedel, 5. *De lenige liefde*. Manteau, Antwerpen, 31.

o, ik weet het niet,
maar besta, wees mooi
zeg: kijk een vogel
en leer me de vogel zien.
zeg: het leven is een brood
om in te bijten en de appels zien rood
van plezier, en nog, en nog, zeg iets.
leer me huilen, en als ik huil
leer me zeggen: het is niets.

29. Vinkenoog, S. (1972). Rolling Stone. *Wonder boven wonder, gedichten 1965-1971*. De Bezige Bij, Amsterdam, 50.

Rolling Stone

ZET JE AAN HET ROLLEN.
Wees niet bang voor grensoverschrijdingen,
je valt van deze wereld heus niet af,
de wereld is rond, je kunt blijven rollen,
ga en ga en ga en ga en zet je ogen wijd open,
want het leven is een reis die je maar één maal kunt maken

30. Marsman, H. (1975). De dierenriem, VI. *Verzamelde gedichten*. Querido, Amsterdam, 149.

De dromen gaan door zijn slaap
als gedrochten door het heelal;
de maan is een dier dat vergaat
in het schaamteloos wolkendal;
en hem, wien het vuur van den geest
met den beet van een schorpioen
door het glad labyrinth van de schors
in het weke der hersenen drong
als gif in een gulzige spons,
breekt bij nacht in een doornenkroon
het zweet der gedachten uit
als een schimmel, een venuskrans.

31. Habakuk II de balker (1970). De steen. *Boerengedichten*. De Bezige Bij, Amsterdam, 85.

De steen

De mens is een steen
Uit een steen gekomen
slaapt hij weer in onder een steen

Ringsteken en radslaan
Doodslaan is zijn opdracht is het niet?

Doorgesneden toont hij als een appel
een rode ster, een hartig centrum
Lust, list, schrift

O de eenzaamheid
van de steen langs de weg
onder de weegbree

32. Campert, R. (1976). Deze middag. *Alle bundels gedichten*. De Bezige Bij, Amsterdam, 100.

Deze middag

Deze middag is een mand
waarin de dagen vallen en vormen maand voor maand,

jaar na jaar, geleefd en nog te leven

De zon is van altijd
en van overal, van het groene gras van Ockenburg
van verleden jaren en van morgen's, overmorgen's witgekalkte dorpen

En de wind, een zacht knippen van de vingers
om de helmplant te herinneren, de tere zandstorm
in geboorte's tuin en dood's Sahara

33. Lodeizen, H. (1963). Hij of zij, 2. *Het innerlijk behang en andere gedichten*. Van Oorschot, Amsterdam, 168.

deze zachtheid
is jou zo eigen, je mond
is een huis geweest waarin
wij woonden, slapend
en biddend.

door jou
heb ik de wereld genomen
heb ik de avond niet gewantrouwd
door jou heb ik geslapen
en geleefd.
door jou ken ik de liefde.

34. Claus, H. (1966). April in Parijs. *Gedichten (1948-1963)*. De Bezige Bij, Contact, Amsterdam, Antwerpen, 119.

de nacht is een vrouw
o honderdduizend lippen
en met de morgen komen twee gelijke treurige Chinezen
in ons wakkerwordend huis
en zeggen ongehoorde zinnen met hun handen
over kastelen of gevangenissen
(zij kijken door de tralies van hun vingers)

35. Vinkenoog, S. (1972). Drugsong. *Wonder boven wonder, gedichten 1965-1971*. De Bezige Bij, Amsterdam, 95.

Drugsong

Tuimel mij binnen in het dwangbuis van de tijd -
het laatste allerschandaligste rechteloze
krampachtig opgekropte stuiptrekken

de poëzie is een beer die zijn adem inhoudt,
een panter in zijn vaart gestopt,
een damhert met het oor te luisteren,
een diepe zuchtende ademhaling

36. Campert, R. (1976). Solo. *Alle bundels gedichten*. De Bezige Bij, Amsterdam, 307.

Solo

lang mee stil staan op deze plek bv.
hier en daar wat struikgewas
een enkele fietser die zijn weg zoekt
of weet laat ik nuchter beschrijven
een paar neonlampen een schutting
met reclame voor benzine en voor de voetbalwedstrijd
de rechte dode weg het verkeersbord en de witte
mijlpaal er trilt iets
iets trilt er de kunst
is op het punt van ontstaan haal diep
adem nu de adem maakt de regel en de regel
beïnvloedt de adem poëzie
is een beest op het punt van de sprong een golf

37. Polet, S. (1977). Poëzie, I. *Gedichten I*. De Bezige Bij, Amsterdam, 138.

Poëzie

Hij had niets te eten.

Daarom at hij niets,
minutenlang.

Armoede is geen erfzonde.
Daarom gaat hij naar huis.

Ruimte genoeg in zijn hoofd, speelsel
en speling genoeg. Hier

hoeft hij niet om adem
te ruziën, de poëzie

is hij zelf, zelf
1 meter sterker, spier-

rijker. Ha,
de poëzie is een mens

-mits vergezeld van een mens.

38. Geel, J.C. van (1973). Vogel. *Enkele gedichten*. Atheneum, Polak & Van Gennep, Amsterdam, 53.

Vogel

De stilte is een brug op vleugels.
Niet dat een vogel vliegt verbaast,
maar dat hij het geruisloos doet.

39. Snoek, P. (1982). Gedicht om de dag mee te beginnen. *Verzamelde gedichten*. Manteau, Antwerpen, 614.

Gedicht om de dag mee te beginnen

Soms hoor ik tussen mijn gewrichten

het kraakbeen knerpen als piepschuim dat breekt,
als dikke sneeuw van vroeger maar dan dun
en hard geworden, scherper, droger en verouderd.

De gloed van het verleden is gekoeld.
Mijn toekomst is een hart vol lege zakken.
Ik heb nog hout voor vele lange winters
maar mijn lichaam heeft de koude aanvaard.

40. Lucebert (1974). V. *Verzamelde gedichten*. De Bezige Bij, Amsterdam, 125.

V

onze vrede is een huid van vlammen
onze vreugde is een huis vol vuur
onze aandacht is kappen op valken
een opengespalkte angel is ons inzicht

onzichtbaar in mijn gezicht
maar in mijn gezicht
eetbaar zijn je ogen
groen smaakt de lente van je iris
en je stem is een aster in de eistille herfst

onze duisternis is een obelisk van regen
onze zachte stem is een warme ruiker
fluisteren en zwijgen maken een etmaal
en zon en maan ons zingend lichaam

41. Hillenius, D. (1971). De woede is een slang. *Tegen het vegetarisme*. Van Oorschot, Amsterdam, 60.

De woede is een slang
die te lang slapen blijft
lijf open laat liggen voor pijn

De woede moet opgeroepen
uit haar kleine mand
zij is een goed vergif
tegen kleinerend verdriet

42. Claus, H. (1966). De zanger. *Gedichten (1948-1963).* De Bezige Bij, Contact, Amsterdam, Antwerpen, 129.

De zanger

Vrij is de zanger niet
Maar vlug en schamper en toppen scherend als een baard.

Vrij is hij niet want zijn doorboord geklater
En zijn sprekend wormhout hangen in zijn mond, huig en tong.
De zanger is zijn lied.
Losgelaten in zijn huid, dit huis,
Groet hij koekoek noch vinkenvanger.

Noch de schuwe spieders in het laagland.

43. Hillenius, D. (1971). De zee is een buik vol embryos. *Tegen het vegetarisme*. Van Oorschot, Amsterdam, 46.

De zee is een buik vol embryos
vol roeivoetigen, kopvoetigen, veelvoetigen,
vol voeten die 't lopen nog leren moeten,
niet verder zijn dan sierlijk bewegen van vinnen
(gedagwuiven is een restje van zwembeginnen)

de zee is moeder van koningen,
van zeepaardjes en malle ponen.
Van wieren is het oudste bewegen van leven
elk gebaar dat we namen gegeven
van strelen, protest, van vechten en vreten
is, buiten ons zelf bestuurd,
in af- en aanstromend water,
hetzelfde van vroeger gebleven.

44. Elburg, J.G. (1975). *Gedichten 1950-1975*. De Bezige Bij, Amsterdam, 126.

Ik zeg: de zon is een druif, nee een meloen
een verse vijg, nee een inktvis een palmkruin.
ik zeg: de zon is een hatende catalaan.
nee een kruik vol wijn, nee een mes of een keisteen.
ik moet snel fotograferen met allebei mijn ogen
nemen met mijn handen, eten met mijn mond.
het zuiden komt op mijn gezicht over:
dikke lippen, oren van brons.
ik heb honger
een warmbloedig zeedier
ik kan veel vrouwen aan.

45. 'De wind is de boom'. Schierbeek, B. (1978). De derde persoon. *Het boek ik; De andere namen; De derde persoon*. De Bezige Bij, Amsterdam, 438.

CHAPTER 5: FEATURE ELICITATION*

5.0 **Feature sets and their comparison**

In metaphor theory, the measure of overlap in meaning between *A*- and *B-term* plays an important role. The interpretation of a metaphor is supposed to be simpler when *A*- and *B-term* show a certain resemblance. How should meaning correspondence be measured? Several methods are available (cf. Miller & Johnson-Laird 1976: 237-260). One way is to compare association-hierarchies in the form of feature sets. If two terms share many word associations, they are often related in meaning. Although this is not necessarily true, and theoretically not well understood, the method may have certain validity as a means of measurement. Thus, the shared set will be used to investigate the meaning relatedness of *A*- and *B-term*.

What should be considered a feature? In the strict sense, the features of a doctor are, for instance, 'gives medical treatment' and 'cures'. 'Nurse' in this sense is not considered a feature, although it may be one of the first associations on 'doctor'. Nevertheless, 'nurse' as an association on 'doctor' may be important to interpret a metaphor, such as 'love is a doctor'. In that case, the message may not only be that love cures some existential pain - as the doctor does. However, it may also say that love looks after those who are hurt - as the nurse does. Put differently, the interpretation of a metaphor is not necessarily only dependent on the features in the strict sense, but also on the close associations. In this study, therefore, features are considered to be all those words and associations that occur in response to a given *A*- or *B-term*.

> Disentangling the effects of linguistic and nonlinguistic associations is a persistent headache in psycholinguistic experimentation; it would be absurd to suggest that arbitrary responses are of no importance or that linguistic analysis can somehow explain them away. (...) both linguistic and nonlinguistic mechanisms are required (...). (Miller & Johnson-Laird 1976: 249-250)

Section 5.1 argues that notions in the theory of literature, such as equivalence, opposition, paradigm, and isotope, can be restated as problems referring to set theory. Section 5.2 presents methods and research that is concerned with association-hierarchies in the form of feature sets. In Section 5.3, new propositions are made with respect to feature salience and the measurement of the shared set.

In Section 5.4, the metaphor models presented in Chapters 1 up to 3 are studied in a feature elicitation experiment, utilizing a *single term*, an *ex-*

---

* Notes are on the pages 205-206; an Appendix with statistics is on 207-225.
Levelt is thanked for the first paragraph of 5.0.

pression, and a *context* condition. Further, an experiment on interpretation is conducted, as well as a rating experiment on similarity and figurativeness. Section 5.5 and 5.6 end this Chapter with conclusions for psychology and the theory of literature.

5.1 Theory of literature: Equivalence as similarity, opposition as dissimilarity

What happens when we read a word? As Riffaterre (1980: 141) puts it, 'beneath the words there is nothing but more words'. In the Sections 1.3, 2.3 and 3.3, reading a word was modeled as the *encoding* of a word, indicating that for every word a feature set of word associations is activated. In the comparison view, the meaning of a word is unchangeable; it is independent of context. For the sake of argument, this theoretical element was also attributed to the anomaly theory, although certain anomaly theorists allow context effects. When it comes to the encoding of words, Lotman could be viewed as such an 'anomalist who considers context effects'. Lotman affirmed that 'poetry consists of words', and reading a (literary) text thus involves the encoding of the ensuing words (the syntagma) by generating the formal and semantic features.

> Poetry consists of words. Nothing could be more obvious. Nonetheless, taken by itself this assertion is capable of generating misunderstandings. The word in poetry is a word from the natural language, a unit of the lexicon, which can be found in a dictionary. Nonetheless it is not equivalent to itself. It is precisely this similarity, this coincidence of the poetic word with the "dictionary word" of a given language that makes the difference between these units so palpable; units that first draw apart and then near, but which are separated and contrasted. (Lotman 1976: 84)

Lotman took a kind of anomaly point of view, in emphasizing that the encoding of a word in poetry is twofold. On the one hand, it is the activation of the dictionary meanings. On the other, it is the activation of nondictionary meanings, primed by the poetical context. In other words, more or other (semantic) features are activated when a word is read in a poetical text. For that matter, Lotman's statements suggest that in *single term* conditions, *A-* and *B-term* only activate the dictionary meanings. In the context of the poem, however, they would activate other (probably figurative) meanings as well.

> (...) a poem as an integral language resembles an *entire* natural language but not a part of it. The fact that the number of words in this language is reckoned in tens or hundreds and not hundreds of

> thousands changes the weight of the word as a meaningful segment of the text. The word in poetry is "larger" than this same word in a general language text. It is easily noted that the more concise the text, the weightier the word and the larger the portion of the universe it designates. (Lotman 1976: 84)

Here, Lotman stated that a word in poetry bears more meaning than in an ordinary text. Supposedly, then, the size of the feature set is larger when a word is part of a poetical text than when it is part of a paper article. Lotman also stated that a word in a short text covers more meanings than the same word in a larger text. Thus, readers should show greater associative response (should elicit more features) for short texts than for long. Thirdly, the feature set is probably larger for a word surrounded by text, than for that word in isolation. Ergo, the number of features that readers elicit for a word in a poem should be higher than the number they elicit in a *single term* condition. In Section 5.4, the last two assumptions are tested.

Lotman's anomaly view is further illustrated, in speaking of the 'conflict' and 'tension' between the dictionary (literal) meaning and the meanings compelled by the poem (the figurative meaning):

> Words, while obtaining special meanings in the poetic structure, also preserve their own dictionary meanings. Conflict, tension between these two types of meanings, is all the more palpable in that they are expressed in the text by a single sign, a given word. (Lotman 1976: 85)

Whereas Lotman showed an anomaly point of view where it concerns the encoding of words, his opinion tends to be more interaction-like, when it comes to the comparison of words, 'the concept of parallelism'.

> In examining the role of repetition, the concept of parallellism should also be considered. It has frequently been scrutinized in connection with the principles of poetics. A. N. Veselovsky indicated the dialectical nature of parallelism in art:
>
> It is not a matter of the identification of human life with nature nor of a comparison that presupposes an awareness of the separateness of the compared objects, but of comparison in terms of some feature of an activity.
>
> Thus, in parallelism, analogy rather than identity or separateness is stressed. (Lotman 1976: 88)

Lotman's ideas on parallelism could easily be applied to the functioning of metaphors, resulting in the interaction position that features of two terms - remote at first - are aligned by similarity through the creation of relations.

> These properties of parallelism may be defined in the following fashion: parallelism is a binomial wherein one of its members is known through a second one that functions as an analogue in relation to the first. This second member is neither identical to the first nor is it separate from it. It is in a state of analogy having those common features which are isolated in the recognition of the first member. Recalling that the first and second members are not identical, we equate them in some particular aspect and consider the first in terms of the properties and behavior of the second member of the parallel. (Lotman 1976: 88)

For Lotman, parallelism by creating relations in a metaphor is only one instance of a general tendency in poetry, which is to create similarity and contrast through analogical reasoning. The comparison of the 'object-member' (the *A-term*) with the 'model-member' (the *B-term*) and the resulting transfer of meaning are a matter of analogical reasoning. Moreover, analogy is fundamental to the creation of similarity in poetry:

> Where we are dealing with parallelism on the level of words and word combinations, a trope relationship arises between the object-member and the model-member, since so-called "transferred meaning" is the establishment of analogy between two concepts. This is the source of that "imagery" traditionally considered a basic property of poetry, but which, as we have seen, constitutes only a manifestation of a more general regularity in a comparatively limited sphere. In fact, it might be said that poetry is a structure all of whose elements on all levels are in a state of mutual parallelism and which, consequently, bear a particular semantic load. (Lotman 1976: 89)

The question could be raised how this parallelism is captured in the study of reader interpretation. According to Lotman (1976: 18) 'words are the most widespread type of conventional sign', and 'in the semiotic sciences, language is defined as a mechanism of sign communication serving the goals of storage and transmission of information' (Lotman 1976: 17). In other words, semiotics involves the study of language storage in memory and the psycholinguistic study of information processing. How did Lotman perform this task in analyzing a real example, and what can methodologically be learned from that? On Mayakovsky's poem 'And could you?' Lotman stated:

> Insofar as the text is immediately assigned the opposition "I-you," the interpretation of this opposition presents itself for consideration:
>
> | I | - | You |
> |---|---|---|
> | poetry | | everyday reality |
> | vividness | | banality |
>
> Mayakovsky's text first evokes this organizational system of the text's "world of words" in the reader's memory and then rejects it. In the first place, the entire verb system that links the noun-concepts in the poetic picture points not to the delimitation of the "mundane" and "poetic" semantic fields, but rather to their fusion. (Lotman 1976: 87)

What Lotman tentatively formulated here is the process that the words 'I' and 'You' activate feature sets ('world of words', in this case, the 'mundane' and the 'poetic'), which are manifested in the form of word associations ('poetry, vividness' and 'everyday reality, banality'). The feature sets of the 'mundane' and the 'poetic' are compared, and at first, they are in opposition (show many distinctive features). This is the representation of the first stage in the process. However, the 'verb system' forces the reader to connect these different feature sets, so that a second stage is entered. Here, the two sets are fused, which leads to an interpretation. Probably, Lotman would argue that this interpretation is established through analogical reasoning:

> (...) a single poetic phenomenon may manifest first its integrating and then its relational associations in relation to various units. On the other hand, both of these types are mutually alternative and, consequently, form a certain diversity of association. At the same time both together are alternative to linguistic associations insofar as they form a meaning according to a different synthetic principle and are not analyzable into the mechanical sum of meaningful units. By virtue of the fact that this series of meanings does not abolish linguistic meanings but co-exists with them forming a mutually correlated pair, we are dealing here with an increase of diversity. (Lotman 1976: 100)

Lotman stated that the activated feature sets are compared on the level of linguistic meaning (the literal meaning). The number of shared features between the feature sets is incremented by creating relations, thus raising the perceived similarity. Here, Lotman actually advocated an interaction point of view. Whether similarity can be treated as the function of a mechanical sum of (shared) meaningful units (features) will be tested in Section 5.4.

> Thus, for [a] word (...), a particular sequence of (...) phonemes and a certain morpho-grammatical structure constitute its expression, and

> the lexical, historical, cultural, and other meanings, its content. (Lotman, 1976: 17)

This content is recorded in Section 5.4, as feature lists in an association production task (feature elicitation task). Subjects are asked to generate the lexical, historical, cultural, and other meanings for the words in the metaphoric, literal and anomalous expressions.

The previous paragraphs clarified that the traditional practice of interpretation is based on the creation of feature sets through the association of words. To connect these sets, a *comparison* takes place that detects parallelism (similarity) and opposition (dissimilarity):

> Le langage poétique connaît un procédé élémentaire: le rapprochement de deux unités. Les variantes sémantiques de ce procédé sont: le parallélisme, la comparaison (cas particulier du parallélisme), la métamorphose (parallélisme projeté dans le temps), la métaphore (parallélisme réduit à un point). (Jakobson 1973: 21)

> In poetry not only the phonological sequence but in the same way any sequence of semantic units strives to build an equation. Similarity superimposed on contiguity imparts to poetry its throughgoing symbolic, multiplex, polysemantic essence which is beautifully suggested by Goethe's "Alles Vergängliche ist nur ein Gleichnis" (Anything transcient is but a likeness). Said more technically, anything sequent is a simile. In poetry where similarity is superinduced upon contiguity, any metonymy is slightly metaphorical and any metaphor has a metonymical tint. (Jakobson 1981: 42)

As indicated by Jakobson, the ultimate poetic tool is the mutual connection of units through parallelism, which may be found in comparisons and metaphors. This parallelism - 'the units that strive to build an equation' - is accomplished when 'similarity is superinduced upon contiguity'. This idea goes way back to the theoretical writings of the poet Hopkins (1966, originally 1865), stating that metaphors are based on perceiving parallels and contrasts, likeness and unlikeness, in a word, on the perception of similarity and dissimilarity between two units.

> The artificial part of poetry, perhaps we shall be right to say all artifice, reduces to the principle of parallelism. The structure of poetry is that of continuous parallelism, ranging from the technical so-called Parallelism of Hebrew poetry and the antiphons of Church music up to the intricacy of Greek or Italian or English verse. But parallelism is of two kinds necessarily - where the opposition is clearly marked, and where it is transitional rather or chromatic. Only the first kind, that of marked parallelism, is concerned with the structure of verse - in

> rhythm, the recurrence of a certain sequence of syllables, in metre, the recurrence of a certain sequence of rhythm, in alliteration, in assonance and in rhyme. Now the force of this recurrence is to beget a recurrence or parallelism answering to it in the words of thought and, speaking roughly and rather for the tendency than the invariable result, the more marked parallelism in structure whether of elaboration or of emphasis begets more marked parallelism in the words and sense. (...) To the marked or abrupt kind of parallelism belong metaphor, simile, parable, and so on, where the effect is sought in likeness of things, and antithesis, contrast, and so on, where it is sought in unlikeness. (Hopkins 1966: 84-85; Jakobson 1981: 39)

How then, does the theory of literature account for this perception of (dis)similarity in poetry? What model is offered to explain how parallelism functions in (poetic) language, and how should it be measured?

> What is the empirical linguistic criterion of the poetic function? In particular, what is the indispensable feature inherent in any piece of poetry? To answer this question we must recall the two basic modes of arrangement used in verbal behavior, *selection* and *combination*. If "child" is the topic of the message, the speaker selects one among the extant, more or less similar nouns like child, kid, youngster, tot, all of them equivalent in a certain respect, and then, to comment on this topic, he may select one of the semantically cognate verbs - sleeps, dozes, nods, naps. Both chosen words combine in the speech chain. The selection is produced on the basis of equivalence, similarity and dissimilarity, synonymy and antonymy, while the combination, the build-up of the sequence, is based on contiguity. *The poetic function projects the principle of equivalence from the axis of selection into the axis of combination.* (Jakobson 1981: 27)

Central to the mechanism Jakobson proposes are the notions that selection is based on (dis)similarity, and combination on contiguity. Underlying the idea of selection is that a noun, such as 'child', is chosen from a noun set that refers to children (the paradigm). This implies that the sender of a message activates nouns that refer to children, before selecting the appropriate one. Reversely, reading the word 'child' probably activates similar nouns, such as 'kid', 'youngster' and 'tot'. Therefore, the paradigm should be represented in the feature set of 'child'. The paradigm words are thought to maintain a high degree of similarity. In other words, each word in the 'child'-set shares many features, thus forming a noun-set with high inner-group similarity. Indeed, in this connection the same mechanism applies to the set of verbs, such as 'sleeps', 'dozes', 'nods', 'naps'. This verb-set also has a high similarity within the group, and each of its members has a high probability to be evoked by any other member of the set.

Furthermore, Jakobson assumes that if two members from different sets (e.g., 'child' and 'dozes') are combined in normal speech, these words should be cognate. In order to be combined, they should have a great deal in common. It could be easily argued, that Jakobson's notion of combination is also based on (approximate) similarity. In ordinary speech, only those verbs which apply to 'child' are likely of being a feature of a child. Put differently, 'sleeps' has a high probability of being incorporated in the feature set of 'child'. In turn, 'child' has a high probability of being in the feature set of 'sleeps', at least in contrast with other verb-sets, such as 'soldiering' or 'to preside at'. Neighbouring words in a sentence, therefore, will probably also have a high proximity in feature sets. They should share many features, or at least incorporate each other in their sets. Contiguity in Jakobson's sense, then, is also based on similarity, particularly, between words from different linguistic categories, such as nouns and verbs. Barthes would agree on this:

> *Similarity and dissimilarity; difference:* The system constitutes the second axis of the language. Saussure has seen it in the shape of a series of associative fields (...). Each field is a store of potential terms (since only one of them is actualized in the present discourse). (...) The terms of the field (or paradigm) must at the same time be similar and dissimilar, include a common and a variable element (...). (Barthes 1970: 71-72)

From this quotation, it readily follows that neighbouring words in normal speech should show a high similarity within and between the members of the different paradigms (e.g., a noun-set and a verb-set). Poetic language, however, in connecting unexpected, improbable members of each other's feature sets, should show high similarity within the paradigm, and low (or less) similarity between paradigms.

To relate the above to a real example, Section 5.4 describes an experiment in which literary students associated on the single words of the expressions from Table 4.1 (Chapter 4). One metaphor ('the child is a moon') contained the word 'child' as the topic of the message. To illustrate how Jakobson's and Barthes' assumptions can be controlled, the data obtained for this metaphor are contrasted with those for its literal counterpart ('the child is a human being') in the next paragraphs.

From the quotations above, it can be predicted that in the *single term* condition, 'child' activates within-paradigm words, such as 'tot', 'youngster', 'kid'. Perhaps, it will even activate without-paradigm words, such as 'sleeps', and 'dozes', because they are closely related. It is also predicted that the number of shared features between 'child' and 'human being' is higher than between 'child' and 'moon', because the first pair lies within the same paradigm, whereas the second lies outside (essentially, Jakobson maintains a comparison view, here). Moreover, 'child' and 'human being' are probably included in each other's feature set, so that the shared feature set includes features of

Table 5.0: The frequencies of features mentioned in a sample of 20 subjects, for the words 'child', 'moon', and 'human being', in the *single term* condition of the feature elicitation experiment. From these lists, feature overlap was calculated. Frequencies preceed the features. Frequencies smaller than 2 are only tabulated for the shared sets. Double numbers in front of the shared features indicate the frequency of occurrence in response to the *A*- and *B-term*, respectively. Translation from the Dutch.

feature set **child**	*feature set* **human being**	*feature set* **moon**	*shared set* **child - human being**	*shared set* **child - moon**
13 small	12 animal	14 night	1 1 birth	1 2 blue
9 young	9 man	9 full	1 2 clothing	1 1 love
8 parents	9 woman	6 shine	1 1 colors	1 1 sleep
7 baby	7 walk	6 stars	1 1 eyes	13 1 small
7 mother	7 war	6 sun	1 1 friends	
7 play	6 live	6 werewolf	1 5 I	
6 school	5 earth	6 white	1 1 ignorant	
5 sweet	5 I	5 crater	1 1 in love	
5 tot	5 monkey	5 man in the moon	1 6 live	
4 father	4 child	5 round	1 2 love	
3 girl	4 talk	5 saint nicolas	1 1 much	
3 later	4 think	4 beautiful	2 4 talk	
3 nice	3 bad	4 heaven	1 7 walk	
2 annoying	3 beast	4 landing		
2 baby sitting	3 civilization	4 romantic		
2 boy	3 evolution	4 yellow		
2 care-free	3 feet	3 far		
2 cheerful	3 many	3 light		
2 cry	3 peace	3 planet		
2 fun	3 world	3 rocket		
2 half-grown	2 adam	3 space		
2 human	2 body	3 star		
2 innocent	2 communication	3 universe		
2 learn	2 clothing	2 armstrong		
2 nipper	2 dumb	2 blue		
2 outside	2 god	2 crescent		
2 playful	2 house	2 dark		
2 play ground	2 humain	2 earth		
2 pregnant	2 love	2 ebb		
2 read	2 robot	2 flood		
2 talk	2 society	2 full		
	2 sort	2 half		
	2 up-right	2 high		
		2 satellite		
		2 space traveling		
		2 teath		

high-frequent mentioning (the readily available ones). The shared set between 'child' and 'moon' should not necessarily include high-frequent features, and should thus show a larger percentage of low-frequent features in the overlap. In Table 5.0, the feature sets are printed for 'child', 'moon' and 'human being', along with the two sets of shared features.

Table 5.0 shows that certain features - apparently belonging to the paradigm of 'child' - are present in the 'child' set (baby, tot, nipper, kid). Indeed, some of these features are even the higher-frequent ones (those with a frequency, higher than 1). 'Baby' is mentioned 7 times, 'tot' 5 times, and 'nipper' twice. The paradigms of 'child' and 'to sleep', however, are but weakly cognate. Only once 'sleep' occurred as a feature of 'child'.

Being one of the readily activated, higher-frequent features, 'child' was mentioned 4 times in response to 'human being', whereas 'human being' was mentioned twice in response to 'child'. Contrariwise, 'moon' did not occur in the feature set of 'child', nor did 'child' in the 'moon' set. Thus, the assumption is illustrated that 'child' and 'moon' do not belong to cognate paradigms. With reference to the Sections on category membership in Chapter 4, 'child' is a good example of instance dominance (the availability of the exemplar, given the category-name), and 'human being' of category dominance (the availability of the category-name, given the exemplar). Table 5.0 illustrates also that the feature overlap between the related terms (child - human being) is larger (13 features), than the overlap between terms from the metaphor (child - moon: 4 features).

The prediction that the feature overlap in 'normal language' contains a larger percentage of high-frequent features than the overlap in poetic language is not corroborated. Suppose that 'the child is a human being - as a literal expression - is an example of normal language, and that 'the child is a moon' - being a metaphor - represents poetic language. 'Child' and 'human being' shared 13 features, of which six features received a frequency higher than one. Thus, the portion of high-frequent features in the feature overlap between 'child - human being' was 6 / 13 = 46%. For 'child - moon', four shared features were found, and two times, these features were higher-frequent. Thus, 2 / 4 = 50% of the overlap between 'child - moon' consisted of high-frequent features, which is about the same as for 'child - human being'.

Of course, illustrating assumptions by analyzing one example is a naive way of arguing. Moreover, taking together the features from different subjects without considering individual differences has some disadvantages, which are discussed in Section 5.2. Therefore, Section 5.4 demonstrates a more rigorous approach to calculating feature overlap, thereby analyzing feature sets and shared sets within subjects. Nonetheless, the analysis of Table 5.0 shows that traditional assumptions in the theory of literature can be quantified and controlled for fairly precisely and that predictions can be derived that may hold for more than one example.

Thus far, it was pointed out that the primary semantic equivalences within paradigms, and between the selected members from different paradigms in the sequence of the sentence are based on similarity. However, this mechanism also applies to the forming of text structures through the connection of isotopes, which are defined as primary semantic equivalences within a text. Isotopes also are thought to be based on equivalence, or on a connective term (Greimas 1966: 71). This may be either an overall category or a shared feature. However, as argued in Chapter 1, an overall category *is* a shared feature. In the theory of literature, an isotope is a connective chain of words across a text, which shows likeness among its members, either formal or semantic. In other words, these words in the chain have a high degree of similarity, thus, share with each other more, and more important, features than with any other word in the text. What happens, when two of these isotopes are compared?

> Les deux isotopies sont reliées entre elles par le *terme connecteur* commun. Dans les cas les plus simples (calembours, << esprit des

mots >>, etc.), l'identité, ou même la simple ressemblance du formant, suffit pour connecter le deux isotopies (...). Le plaisir << spirituel >> réside dans la découverte de deux isotopies différentes à l'intérieur d'un récit supposé homogène.

On voit, par conséquent, que le << bon mot >>, considéré comme genre littéraire, élève au niveau de la conscience les variations des isotopies du discours, variations qu'on fait semblant de camoufler, en même temps, par la présence du terme connecteur (...).

La confrontation de deux isotopies met en opposition (...) non plus seulement deux séquences possédent chacune un caractère isotope: ces tranches du discours sont considérées, du point de vue de leur contenu, comme représentatives de *mentalités* hétérogènes. (Greimas 1966: 71)

Greimas stated that two isotopes are connected by a common connective term, by identity or by simple resemblance. These are all cases of the same underlying idea that through the comparison of two word-chains, similarity between the words of these chains is discovered. In line with Greimas, the perceived similarity may be the result of a common connective term, which can be seen as a common category for (some of) the words in the two isotopes. The similarity may also be based on identity, in which case the two isotopes share the same words. Similarity may also be found in resemblance, which means that the feature sets of the words in the first isotope are compared with the feature sets of the words in the other isotope, between which shared features may be detected. In contrast, if the isotopes are in opposition ('isotopies met en opposition'), the degree of dissimilarity dominates the perceived similarity. Many words, or features of words, form mismatches between the two sequences.

From a perception point of view, a method should be postulated to investigate the relations of similarity and dissimilarity among signs and words. Greimas defined the 'termes-objets' (words/lexemes) as their 'sèmes', as the properties of the words, or their features, for that matter.

(...) on doit essayer maintenant de déterminer le rôle qui peut être assigné (...) aux termes-objets, dont, au niveau de la perception, nous avons postulé l'existence en même temps que celle de la relation.

Nous avons vu que cette dernière pouvait être analysée en sèmes, que nous avons proposé de considérer comme des propriétés des termes-objets.

(...) Tout cela ne fait que confirmer notre répugnance à l'égard d'une sémantique qui aurait la prétention de décrire la << substance psychique >>. Force nous est donc de rester sur le plan phénoménologique, c'est-à-dire linguistique, et de postuler, avec Russell, que les *qualités* définissent les *choses*, c'est-à-dire que le sème *s* est un des éléments constituant le terme-objet [lexème], et que

> celui-ci, au bout d'une analyse exhaustive, se définit comme la collection des sèmes s_1 , s_2 , s_3 , etc. (Greimas 1966: 27)

To write a semantics for the 'psychic substance' of words, the things ('les choses') should be defined as their qualities (Russell) or properties (which is the same as 'sèmes', according to Greimas). Put differently, the psychological contents of a word are found in the feature sets that result from feature elicitation experiments with a representative sample drawn from a speech community. This may be a dated and almost a behaviorist approach to the psychological contents of word meaning, but it shows how certain assumptions can be put in operation, which would be a gain for semiotics.

On the whole, then, each word evokes a collection (X) of semes, $X = \{s_1 , s_2 , s_3 ,..., s_n\}$, which is a set of features. Such a feature set may be compared with any other set to find the shared features, thus increasing similarity or equivalence within and among sets. Consequently, the features that are not shared by two words (the distinctive features) increase the dissimilarity or opposition. Textual structures (e.g., isotopes) are established by the contrastive weight of the shared and distinctive features of the words in the text. Words with high numbers of shared features are connected into an isotope, and contrasted with other isotopes. Thus, isotopes may be defined by a higher number of shared features within than among sets.

> The interaction between syntactic morphologic and lexical equivalences and discrepancies, the diverse kinds of semantic contiguities, similarities, synonymies and antonymies, finally the different types and functions of allegedly "isolated lines", all such phenomena call for a systematic analysis indispensable for the comprehension and interpretation of the various grammatical contrivances in poetry. Such a crucial linguistic and poetic problem as parallelism can hardly be mastered by a scrutiny automatically restricted to the external form and excluding any discussion of grammatical and lexical meanings. (Jakobson 1981: 90-91)

To comply with the demands put forward by Jakobson, Chapter 4 described how to minimize the interactions between syntactic, morphologic, and lexical equivalences and discrepancies. By means of subject generated feature sets, Chapter 5 proposes a way to analyse systematically the semantic contiguities and similarities between words. A further demand is also incorporated in the tests on the metaphor models. Section 5.4 concerns the conditions under which the semantics of the metaphors ought to be studied: The *single term*, *expression* and *context* condition. The first condition checks the comparison (and anomaly) presumption that context does not change meaning. The second and third test whether metaphors in isolation are differently processed ('the different types and functions') from metaphors in the poetical text.

To recapitulate the Section thus far, the terminology of the theory of literature can be reduced to a convenient set of simple set theoretical notions. The theory of literature reserves a primary role for the perception of similarity (or equivalence). The theorists discussed above usually connect similarity with 'chains of association', 'properties', 'paradigm words', etc. It was deduced that the theory of literature envisions similarity as a function of the number of shared features between two feature sets.

Jakobson's ideas, for example, can be restated as follows: A paradigm is a set of words in the mental lexicon that reaches the highest within-set similarity, compared to words excluded from this set. These within-set words have high probabilities to be mentioned as features of any of the other words within the set. They are the high-frequent features of any word in the set, because they are easily activated by each other. Words selected from a set are put in a sequential order (the syntagma, the combination of words in a sentence) with words from other sets with high within-set similarity. In poetry, the degree of between-set similarity may be less than in ordinary speech.

'Equivalences', 'parallels', 'contiguities' are all tautologies for the notion of similarity between two (or more) words. An isotope is a set of words in a text with a higher within-set similarity than with any other word or word set in the text. Comparison of isotopes may lead to the perception of, again, similar words or identical features of words. Together with the distinctive words, they form the text structure. Henceforth, the tripartite tool 'feature set, similarity, dissimilarity' accounts for the full of intuitions formulated by the theory of literature, thus offering a systematic and quantitative approach to verify if:

> In poetry, any conspicuous similarity is evaluated in respect to similarity and/or dissimilarity in meaning. (Jakobson 1981: 44)

Indeed, the empirical investigation of similarity even drew on semiotics. Referring to a famous study by Tversky (1977), Sonesson argued that the relation between icon and sign may be based on similarity, but that this similarity is not symmetric:

> (...) iconicity cannot motivate a sign, for while similarity is symmetrical and reflexive, the sign is not. Pigments on paper, or carvings in a rock, could stand for a man, but not the reverse; nor will they, in their picture function, stand for themselves. This argument is quoted and accepted by Groupe μ (...); yet, if it is not refuted, only nonsense will remain of iconicity. The error consists in identifying the common sense notion of similarity with the equivalence relation of logic. No doubt, the equivalence relation, as defined in logic, is symmetric and

> reflexive, and thus cannot define any type of sign, since the sign, by definition, must be asymmetric and irreflexive. But to identify similarity with the equivalence relation is to suppose man to live in the world of the natural sciences, when in fact he inhabits a particular sociocultural Lifeworld. *Similarity*, as experienced in this Lifeworld, is actually asymmetric and irreflexive. Indeed, this fact is not only intuitively obvious, but has now been experimentally demonstrated (notably by Rosch 1975 and Tversky 1977). (Sonesson 1996)

Thus, the notion of equivalence is too strict to describe what is meant by the theory of literature. Equivalence is a logical term, demanding symmetry between the compared terms ($A = B$). What the theory of literature envisions is similarity (A is like B), leaving room for dissimilarity and asymmetry.

Tversky (1977) was the first to demonstrate asymmetry in judgements on similarity. Subjects judged systematically that, for instance, 'the portrait resembles the person' was a better comparison than 'the person resembles the portrait'. In the next Section, Tversky's explanation of asymmetry in similarity judgements is thoroughly examined, thereby relating his findings to feature set differences. For now, as an example of similarity research in the theory of literature, one of the scarce but laudable empirical explorations of Jakobson's formal-semantic equivalence theory is discussed.

Wolff's (1977) investigation of Baudelaire's 'Les Chats' did not rely on feature elicitation, but on categorization experiments. Following Miller (1969), Wolff tried to determine the classes of semantic equivalence that compose the poem by asking subjects to categorize those verbs, nouns and adjectives that - according to their opinion - were semantically cognate. The number of subjects that put the same words together was assumed to be an index of the semantic equivalence of these words. The height of their relative correlations was assumed to decide the cluster (or 'context-metaphor') to which they belonged. Although on a descriptive level this approach was highly informative, certain flaws occurred that could have been avoided by a feature elicitation experiment.

Above all, Wolff supposed that subjects were judging the words on *semantic* equivalence, although equivalent spelling, rhyme or syntax could interplay (cf. Chapter 4). No controls were offered on these factors. Furthermore, equivalence matrices are hardly ever symmetric (correlating completely on, and being equal above and below the diagonal). In addition to the cluster analysis, then, arbitrary choices must be made ad hoc, to determine in which cluster a doubtful case should be categorized. Moreover, the equivalence matrix offered by Wolff (1977: 100-101) did not take into account the correlation values above the diagonal, thus presuming that there is no asymmetry in the semantic equivalence of the words in 'Les Chats'. It could be ventured, however, that subjects judge the semantics of e.g., 'sphinx' to be more similar to 'mystique' than 'mystique' to 'sphinx'. Tversky

(1977) offers ample experimental support that subject judgements of equivalence (or similarity) may be liable to such asymmetries.

In addition, it cannot be fixed precisely from an equivalence matrix, whose features distinguish the one cluster from the other, and whose features are shared by the words within the cluster. Usually, such results are arrived at by a heuristic search through the raw material. It is assumed also that the clusters derived from statistical analysis are ecologically valid, thereby assuming that subjects will perceive fewer clusters than there are words or semantic features of the words in the poem.

However, Wolff (1977: 54-55) was not unaware of the problem that lacking knowledge of the semantic features of the words leads to a deadlock in describing the semantic equivalence of words. In rejecting Eco's (1972: 116) loose enumeration of features for different classes, Wolff saw no solution to the problem that feature sets may be infinite and exchangeable. Another problem which Wolff signaled was that the hierarchical models describing the relations between features usually do not fit the data. This is not a pitfall, but rather the very essence of scientific enquiry.

The first problem, on the contrary, is worth attending. Indeed, feature sets may be infinite and interchangeable a priori. However, within a speech community, this is not a requisite. In other words, to know what a word means according to a homogeneous group of people, the accumulated feature sets generated by a random sample is sufficient to estimate the chance that a feature is mentioned. Put differently, statistics provides for the probability of occurrence for each feature in an infinite sample, whereas the sample fences off the feature set. The next Section will discuss how these considerations are put into practice.

5.2 Psychology: Studies on similarity

The issue of equivalence in forming connotative meaning was not only pursued in the theory of literature. The concept of equivalence is also of utmost relevance to the psychological study of similarity.

> Similarity plays a fundamental role in theories of knowledge and behavior. It serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations. Indeed, the concept of similarity is ubiquitous in psychological theory. It underlies the accounts of stimulus and response generalization in learning, it is employed to explain errors in memory and pattern recognition, and it is central to the analysis of connotative meaning. (Tversky 1977)

As argued in the Chapters 1 up to 3, the current theories of metaphor processing attach a more or less central role to feature matching. Comparison theory, for instance, concentrates on the size of the shared feature set (overlap) to differentiate among expression types. Anomaly theory contrasts the shared set with the distinctive features of the *A*- and *B-term*. If too many literal features are distinctive, figurative features are activated to enrich the shared set. Interaction theory claims that when the shared set is outweighed by the distinctive features, relations are created to improve the semantic connections. It should follow that similarity between terms - or 'equivalence' in the theory of literature - depends on the size of the shared set. The more features are shared, the more similarity.

In this respect, certain phenomena are of interest. Tversky has (1977) pointed out that comparisons such as similes and metaphors should not be assumed as symmetric by nature. However, others such as Gleitman, Gleitman, Miller & Ostrin (1996) object to this idea - see *Asymmetry in judgements of similarity and figurativeness?*. Tversky stated that the number of features the *A-term* shares with the *B-term* equals the number of features the *B-term* shares with the *A-term*. Consequently, similarity of the *A-term* with the *B-term* should equal that of the *B*- with the *A-term*. Yet, this is not always the case. Tversky found that subjects preferred phrases such as 'the son looks more like the father' to 'the father looks more like the son' or 'an ellipse resembles a circle' to 'a circle resembles an ellipse'. Although logically invalid, subjects yet judged that the one term was more similar to the other than conversely.

Tversky's explanation of this asymmetry (the focusing hypothesis) ran as follows. The choice of the subject of a comparison (partly) depends on its relative salience. The less-salient (less prominent) term is chosen as the subject or variant (the son). The more salient term is seen as the referent or prototype (the father). Since the less salient term (the subject) is put in focus, its features are weighted more heavily than the features of the referent. Thus, shared as well as distinctive features of the subject are weighted heavier than those of the referent. Since the shared features are equal for both terms, the distinctive features of the subject make the difference in the weighing of both terms. If similarity judgements are connected to the number of shared features in contrast to the number of distinctive features, then the similarity of the comparison from subject to referent is less than that from referent to subject. A comparison from subject to referent puts the heavily-weighted distinctive features of the subject in focus and compares them with the less-weighted distinctive features of the referent. As a result, similarity is reduced more by the distinctive features of the subject than by the distinctive features of the referent. Therefore, a comparison from subject to referent is a comparison from low-to-high similarity (coinciding with low-to-high salience), which is preferred to a comparison from high-to-low similarity.

However, the role of salience is rather obscure. It is a container concept, including notions of intensity, feature frequency, familiarity, good form, and informational content. Others, such as Katz (1982), reckon in associative dominance, typicality, fluency and imaginal distinctiveness. Because of this inclusiveness, salience is as vague as 'foregrounding' in the theory of literature. It is unclear whether the composing elements of salience are competitive or supplementary. Is salience an overall measure of the stimulus, or is it the sum of salience estimates for each feature?

It all comes down to ascribing a weight to (the features of) a stimulus, and most of the concepts assumed to contribute to this weight lack construct validity. Exceptions may be feature set size and feature frequency, because they are readily accessible from the feature sets. Feature set size may be a knowledge index for a stimulus, while feature frequencies may indicate the importance of a feature within this knowledge index. The next Section is devoted to this option.

Salience and feature set size tie for explaining asymmetry

Tversky (1977) asked subjects to choose which comparison between countries they preferred: In an order from low-to-high or in an order from high-to-low salience. Salience of countries was judged beforehand by a norm group. It appeared that 'North Korea is similar to Red China' was preferred to 'Red China is similar to North Korea'. The asymmetry was ascribed to a higher salience for Red China than for North Korea. Identical results were obtained for similarity ratings. Low-to-high salience orders yielded higher similarity ratings than high-to-low orders.

In similarity judgments of geometric figures, salience was defined as 'goodness of form'. A 'good figure' (a simple line drawing) was considered less-salient than a 'bad form' (a more complex line drawing). Tversky found that the preferred order of comparison was from the less-complex to the more-complex figure. This was interpreted again as a preference to a low-to-high salience ordering. Rating similarity for both directions of comparison confirmed that similarity for the less-complex figure compared with the more-complex figure was higher than vice versa. Unfortunately in this case, the salience difference was defined a priori, and not derived from actual judgements.

In comparing block letters, Tversky stipulated that a letter is more salient, when it includes the letter it is compared with (e.g., I → F → E). Indeed, subjects judged that, compared with 'including' letters, 'included' letters led to more 'same'-responses than 'including' letters compared with 'included' letters. However, as in the case of the geometric figures, salience was not judged by the subjects.

In the same way, Tversky maintained that long Morse signals are more salient than short Morse signals. In an analysis of Rothkopf's (1957) data, he found that short signals compared with long signals produced more 'same'-responses than comparisons from long-to-short signals.

For color comparisons, Tversky drew on the Rosch (1975) study, in which subjects preferred a phrase like 'off-red is virtually pure red' to 'pure red is virtually off-red'. Similar results were found for comparisons between horizontal, vertical and diagonal lines and lines under different angles. The variant (e.g., off-red / line under unusual angle) was preferred in the first position, and the prototype (e.g., pure red / horizontal line) in the second. Tversky claimed that the stimuli in first position were least salient.

Tversky observed that, for example, 'tiger' is a more likely associate to 'leopard' than 'leopard' to 'tiger'. For an explanation, he dismissed higher 'response availability' or 'commonality' and insisted on salience differences.

However, the methods to measure all these notions were not well established, and to this extent, they merely remain words. Salience may indicate how conspicuous a stimulus is, and perhaps explain asymmetry. Alternatives such as 'commonality' or 'familiarity', however, may be equally compelling.[1] The only direct measure is a rating of each feature or stimulus on certain dimensions. However, it is equivocal which one. Other direct measures are the feature set sizes of the stimuli, and the frequency with which a feature occurs. An advantage of these measures is that subjects are unaware of what is measured, which is not the case in scaling experiments.

Moreover, feature set size may be related to asymmetry as well. Subjects may prefer a certain order of comparison, not because the focused stimulus is less salient, but because it is feature-poor, whereas the frame of reference is feature-rich. Feature-richness, then, as an index of world knowledge may be a serious candidate for explaining asymmetry.[2]

Hence, the question is whether Tversky's findings should be interpreted from a different perspective. Subjects rated salience in one experiment only (the 'overall salience' of the countries). They did not indicate the salience of each particular feature, and thus, might have estimated the relative importance of the countries by the amount of knowledge they had about them. The unsalient, unimportant countries might have been the countries they knew less about (North Korea, Poland). These countries might have activated the smallest numbers of features. The salient countries (Red China, USSR) coincided with the ones they knew best. In these cases, the feature-poor, unfamiliar, countries were chosen as the focus of comparison, the feature-rich countries as the frame of reference. Asymmetry, then, might be explained by the higher probability for a poor feature-set to share its features with a feature-rich set. A comparison from the feature-rich set to the feature-poor set leaves more distinctive features within the focus of attention than the other way round. Psychologically, similarity for poor-to-rich comparisons is judged higher than from rich-to-poor. Relatively more features intersect in the focus of attention, when the feature-poor stimulus is chosen as the core of comparison.

From this perspective, Tversky's results can be perfectly explained by the number of features that a stimulus activates ('feature-richness'). Complex stimuli, such as 'bad' geometric figures or block letters that include other letters are characterized by having more perceptual features. The 'bad' geometric figures in Figure 3 (Tversky 1977: 335), for example, have more edges, lines, and distance differences than the well-shaped figures. Therefore, they have more perceptual features. For the same reason, the 'including' letters may have more features than the 'included' letters. On top of that, they also have the 'included' letters as features. Tversky failed to explain why short Morse signals are less-salient than long Morse signals. In fact, short signals may have less perceptual features (a dot or a dash less). It may also be argued that 'pure red', a horizontal or vertical line or a diagonal with 45^0 angle are more familiar to the subjects than 'off-red' and lines under different angles, and therefore, evoke more features.

Contrary to Tversky's conception, it may even be argued that a prototypical stimulus is *less*-salient, although it evokes more features. Tversky does not explain when a stimulus or feature is more salient than another. It is quite feasible that salience is a measure of infrequency. Something rare is conspicuous. A low-salient stimulus might be a stimulus with a small set of low-frequent features. A prototypical stimulus, however, is well known (otherwise it would not be a prototype), and is probably not salient, because it has large sets of high-frequent features. In addition, salience may also be related to the number of distinctive features of a stimulus within its category. Something deviant (something that shares few features with other stimuli) is conspicuous. Since a prototypical stimulus shares more features with its category members than a nonprototypical stimulus, it is best 'camouflaged', and probably has the least distinctive features. Considering that the weight of the distinctive features (its 'deviation value') may be an indicator of salience, a prototypical stimulus would be least salient.

Explaining asymmetry as a function of differences in feature set size is more parsimonious than explaining it from salience differences. Salience is a conglomerate of factors and cannot be secured without subject ratings. The frequency of features and feature set size are obtained without intentional awareness of tracing salience.

Tversky arrived at the role of salience by a complex inference. First, stimuli are supposed to activate features. The subject of the comparison is chosen as the focus (Tversky did not explain why), and thus, its (distinctive) features are weighted heavier (why?). Since subjects judge that the similarity of the focused stimulus is lower, the frame of reference is more salient (why?).

Feature set size is a simpler explanation. Stimuli activate features whose set sizes decide about the focus of comparison (not necessarily the subject, cf. Table 4.0, Chapter 4). The feature-poor stimulus is chosen as the focus, because it has a better chance to match its features with the feature-rich, so that the relative number of distinctive features is smaller. In this way, the preference to comparison order is explained, and the concept of separately weighing the focused stimulus can be dropped.

Moreover, Tversky treated salience as an overall property of the stimulus. Since the salience of each feature was unknown, a perhaps interesting specification might have been overlooked. It might be that the shared set is not symmetric, as Tversky presupposed. If a shared feature is low-salient (or low-frequent) in the one set and high-salient (or high-frequent) in the other, asymmetry may be explained from salience (or frequency) differences *within* the shared set (cf. Table 5.0). Salience or frequency of the distinctive features, then, may even be ruled out as explanation for asymmetry.

Several formulae to ascertain similarity

Broadly speaking, three approaches can be distinguished in determining similarity: The calculation of distance, informational uncertainty, and the shared set. One example of each option will be discussed here.

Distance metrics do not employ feature sets. Instead, Tourangeau & Sternberg (1982) presumed a dimensional space in which similar stimuli occupy adjacent regions. The dimensions were established by contrasting antonyms on bipolar rating scales (e.g., noble-ignoble, warlike-peaceful). Following factor analysis, the factor scores on the observed dimensions were treated as coordinates, which - if approximately equally powerful - imply that the vectors of the stimuli are more or less the same. The Euclidian distance d between stimuli A and B, then, is calculated as the root of the sum of squares of differences between the m coordinates (a_i - b_i):

$$d(A,B) = \sqrt{\sum_{i=1}^{m} (a_i - b_i)^2} \quad .$$

In this way, Tourangeau & Sternberg estimated the mutual distances between categories (e.g., 'fish', 'birds', 'world leaders'), as well as the distances between members within a category (e.g., 'shark', 'hawk', 'Castro'). A group of subjects rated the category members on a range of bipolar scales and the obtained factor scores decided about their distance. Another group of subjects rated the same categories on similarity. This resulted in a set of stimuli, belonging to more or less distant categories, and occupying more or less similar positions within their respective categories. From these stimuli, metaphors of the form 'the *A* is the *B* among *C*' were constructed (e.g., 'the shark is the hawk among fish') and a third group of subjects rated the aptness of these metaphors. Tourangeau & Sternberg found out that closeness of within-category positions (shark-hawk) increased aptness, whereas closeness between categories (fish-birds) decreased aptness.

Based on this and other experiments, Tourangeau & Sternberg suggested that when interpreting metaphors, subjects see the factors that apply to one category as analogous to those applying to the other category. Categories should be sufficiently distant to make an apt metaphor. Within categories, however, stimuli should occupy similar positions in the dimensional space.

Tversky pointed out that asymmetry is harder to represent by distance than by feature metrics, although distance metrics that account for asymmetry are found in Krumhansl (1978) and applied in multidimensional scaling techniques. However, an optimal spatial fit of the data is usually achieved by a host of transformations; the remaining stress is an effect of the technique - not of the data - and the resulting model is interpreted fairly intuitively. Moreover, distance is a cumbersome measure of similarity, because although stimuli may be close on several dimensions, they may not look alike or have a similar function. For instance, 'death' and 'guilt' are closely related, yet not similar. In other words, distance metrics are biased towards identifying analogy, whereas feature metrics are biased towards similarity. Since analogy is a core notion of interaction theory, distance metrics favor the interaction view, whereas feature metrics favor comparison and anomaly assumptions.

Malgady & Johnson (1980) expressed the influence of salience on similarity in terms of the average *information value* of a stimulus. The concept of uncertainty is concerned with redundancy of information. The information of an expression such as 'the child is a human being' is redundant in contrast with 'the child is a moon'. The first adds nothing to what is already known, whereas the second provides new information. The higher the redundancy of a stimulus, the less its salience. In other words, the information formulae can be employed for estimating salience and surprise effects.

Malgady & Johnson defined the salience of a stimulus in terms of feature probability by way of the common information formula:

$$U(X) = -\sum_{i=1}^{n} p(x_i) \cdot \log_2 p(x_i) \qquad \text{(Garner 1962),}$$

where X = set of features $\{x_1, \dots, x_n\}$
U = informational uncertainty in the feature set X
$p(x_i)$ = probability of the i-th feature in X
n = number of features in X.

To normalize the differences in feature set sizes, a summary measure of salience $f(X)$ was calculated as:[3]

$$f(X) = \frac{U(X)}{\log_2 n} \qquad ,$$

where salience is highest, if $f(X) = 1$,
and salience is lowest, if $f(X) = 0$.

The numerator $U(X)$ is sensitive to the probability distribution of the features, whereas the denumerator $\log_2 n$ is not. As a consequence, salience decreases as $U(X)$ becomes less. Put differently, for fixed n, the denumerator $\log_2 n$ is constant, so that the quotient is proportional to $U(X)$. Moreover, the probability of a feature is defined as the relative frequency of a feature.

Malgady & Johnson used this calculus to establish the salience of terms in metaphors and similes. A group of subjects produced features for all nouns in 12 pairs (= 24 nouns), originating from literary metaphors. This resulted in open-ended feature sets, from which the informational uncertainty was assessed for all nouns. Classified as high or low in salience, the two nouns were used in two metaphors: One in the order from low-to-high salience (*A* is *B*), the other in the order from high-to-low salience (*B* is *A*).

A second group of subjects rated the similarity of the metaphors in both orders (*A* is *B*, *B* is *A*) from 'the first term to the second'. Malgady & Johnson confirmed Tversky's focusing and asymmetry hypothesis. In judgements of similarity, the order of low-to-high salience was preferred in 70% of the cases. About 73% of the subjects reported that the metaphors were understood better, when the *A-term* was less-salient than the *B-term*. The strong connection between similarity and interpretability was also found by Johnson & Malgady (1979).

A second experiment was conducted to study context effects on judgements of similarity and figurativeness in similes (*A* is like *B*), which were constructed from six selected noun pairs of experiment 1. The most salient noun was used as the *B-term*. The similes were presented in six different context conditions, either including adjectives and verbs that supposedly added a feature to the *A-term* set, to the *B-term* set, or that enriched the shared set. For instance, the adjective 'red' in 'Red blood pours like rain'[4] was assumed to add a feature to 'blood', not to 'rain'.[5] The verbs and adjectives were judged by a norm group for their 'adding value'.

One group of 25 subjects compared the six versions of every simile on similarity between the terms, without instructions about the direction of comparison. A second group judged the 'figurativeness' of the similes. 'Figurativeness' was defined implicitly by contrasting a simile with a literal expression, and with an anomaly. Scale values for the six simile versions were calculated in every condition with Thurstone's case V model (Thurstone 1927).

The context conditions could be ranked by the obtained scale values. The rank order of the conditions for *similarity* judgements confirmed Tversky's contrast model that subjects focused on the nonsalient *A-term*. Contexts that cued the shared features ($X \cap Y$) had a positive effect on judgements of similarity, followed by those cueing a distinctive feature of the nonsalient *A-term* (X - Y), followed by contexts cueing a distinctive feature of the salient *B-term* (Y - X). For judgements on *figurativeness*, the focusing effect reversed, because subjects focused on the salient *B-term*: $(X \cap Y) > (Y - X) > (X - Y)$. Malgady & Johnson concluded that *A*- and *B-term* in metaphor and simile are judged differently for their contribution to similarity and figurativeness, the difference lying in the switch of focus.

Until now, Tversky (1977) has only been discussed with regard to similarity ratings. However, Tversky also supposed that similarity is a function of the contrast between *shared* and *distinctive features*. Note that the *A-term* activates feature set X, the *B-term* set Y, and that u is a feature in X and Y. The calculus for shared features looks as follows:

$$f(X \cap Y) = \sum u_X \cdot u_Y / N_u \,,$$

where $f(X \cap Y)$ denotes a salience measure for the sum of all features u in the shared set $X \cap Y$, and u_X and u_Y denote the frequency of occurrence of feature u in response to stimulus A and B. The denumerator N_u denotes the number of stimuli that share feature u. In simpler terms, the shared set equals the total number of common features, including their frequency of occurrence, divided by the number of stimuli in the comparison. Note, however, that by definition the salience measure $f(X \cap Y)$ is equated with feature frequency (frequency of occurrence), which - despite Tversky's objection - may be seen as a stable indicator of response availability.

Distinctive features are calculated by $f(X - Y) + f(Y - X) = \sum a_X + \sum b_Y$, which is a salience measure f for the features that belong to X, but not to Y, plus the salience of features that belong to Y, but not to X. The sum equals the total of distinctive features a in response to stimulus A, plus the total of the distinctive features b in response to stimulus B. Thus, the salience of a stimulus is defined as the summed frequencies of all its features, where distinctive features are calculated as the total number of features in a set (including frequencies), minus the features (including frequencies) shared by another stimulus in the comparison.

In other words, the calculus for the number of shared features replaces salience in terms of direct ratings by salience in terms of frequency of occurrence of features. The calculus for the distinctive features makes salience dependent upon the total number of features (including frequencies) elicited by the subjects.

Tversky's calculus for the shared set offers appealing as well as disadvantageous aspects. The sum of products of the frequency of feature u in response to A with the frequency of u in response to B, ($\sum u_X \cdot u_Y$), assigns little weight to a feature with a low probability, which is a merit. For example, suppose that in a data set $u_Y = 0$, then u does not add to the shared set. However, the frequency u_Y in an infinite sample may be greater than zero (feature u may be occasionally mentioned). If $u_Y = .001$, then u does add to the shared set, namely, with the value of .001 times u_X . In other words, the value of u_Y is taken into account, but infrequent features are not weighted too much. The denumerator N_u in the formula for $f(X \cap Y)$ denotes the number of stimuli that share feature u. As a consequence, rare (and supposedly salient) features in the shared set are assumed to increase similarity more than frequent ones. However, the premise that rare features have greater importance for the shared set has no empirical support.

Two problems arise, then, when applying the calculus to feature elicitation data. One concerns the loss of information, insidious in multiplying and dividing the feature frequencies, the other concerns statistics. Tversky stressed that the calculus for shared features is only applicable to symmetric comparisons (*A* is *B*). The product in the numerator is an overall measure of the frequency of a feature in the shared set. To determine the value of the shared set $f(X \cap Y)$, the summed product is divided by N_u , which makes $f(X \cap Y)$ an overall value that cannot distinguish the frequencies in X and Y of the shared features. Consequently, asymmetry cannot be investigated as a function of frequency differences of features *within* the shared set.

Moreover, this type of data analysis induces a statistical problem. Feature frequency is assessed by the number of times a feature is mentioned by a group of subjects. In other words, feature sets accumulate across subjects, so that the variance within subjects is neglected (this critique is also applicable to Tourangeau & Sternberg and Malgady & Johnson).

Of course, feature frequency can hardly be obtained otherwise, because subjects usually do not mention a feature more than once. On the other hand, inaccuracies may be introduced to the shared set. Feature u_X may be mentioned by subject 1, whereas u_Y is mentioned by subject 2. If the sets are joined across subjects, the shared set $X \cap Y$ contains feature u. However, for each individual subject 1 and 2, the shared set is empty. In other words, when variance within subjects is respected, less feature overlap is found than when the sets are accumulated, which may actually exaggerate the size of the shared sets.

Notwithstanding the risk of data-sets with many zero-values, analysis of feature elicitation data should respect differences between subjects. The gain is that feature u_X mentioned by subject 1 *cannot* be compared with feature u_Y from subject 2, so that shared features refer only to individual subjects. The loss is that frequency differences within the shared set will hardly ever occur.

Hitherto, the studies discussed have been rather dated; however, more recent studies on similarity come up against the same problems on salience and weight of features (e.g., Goldstone, Medin & Gentner 1991). Recent studies on salience and similarity tend to take Tversky (1977) as a starting point. However, as Ortony put it:

> The issue of whether or not salience of an attribute is independent of the object that possesses it is an unsettled empirical question. Yet, it is not at all easy to test. It is difficult to distinguish between the absolute amount of salience an attribute contributes to an object and confounding variables such as the relative amount it contributes and the amount of knowledge that subjects have about the objects.
>
> (...) (presumably highly correlated) measures, such as frequency or order of mention in an elicitation task, might do just as well. (Ortony 1979)

As indicated, various approaches are available, so that the most suitable to the metaphor models should be selected. Since all models rely on activation and comparison of feature sets, Tversky's approach seems most promising. The adjustments are, however, that (1) the data analysis will respect subject variability, (2) salience is conceived as feature frequency, and (3) the frequencies of shared features may differ in *X* and *Y* and may explain asymmetry as well. Section 5.3 proposes a calculus that meets these demands, thereby giving account of the definition of features, and the criteria for identifying shared features.

5.3 Three definitions for features, their formal and semantic equality, resulting in three measures of feature overlap

Similarity judgements may sometimes be asymmetric (Tversky 1977). Subjects judge that stimulus *A* looks more like stimulus *B* than inversely. A major factor in materializing this phenomenon is supposed to be the salience difference between *A*- and *B-term*. Section 5.2 argued that salience is a problematic concept in comparison with feature frequency and set size. Asymmetry, then, is supposedly caused by the term with a feature-poor set. This term is preferred as the focus of comparison, so that relatively more features are covered by the feature-rich set.

Modifying Tversky's approach, the notion of feature salience is dropped, and the simpler frequency of occurrence is maintained. The following dependent variables are investigated: The weighted feature set size (W_X and W_Y - see Chapter 1), and feature frequency differences within the shared set. W_X and W_Y are defined as the sum of weights of all the features evoked by the *A*-

and *B-term*, the weights of which are determined by their frequency of occurence.

The weighted shared set size (W_S - see Chapter 1) is the sum of the weights $w_S(u)$ of all features u in the shared set; weight being the frequency with which a feature occurs. However, 'red' may be mentioned once as a feature of 'love' and twice as a feature of 'rose'. To allow for frequency differences within the shared set, the weighted contribution of X in $X \cap Y$ is established by:

$$_XW_S = \sum_{u \varepsilon S} w_X(u) \ .$$

In the same way, the weighted contribution of Y in $X \cap Y$ is given by:

$$_YW_S = \sum_{u \varepsilon S} w_Y(u) \ .$$

The disadvantage of this approach is that the data set is now considered as the population. The probability that a feature occurs in an infinite sample is not considered. In other words, infrequently reported features receive a relatively high weight.

In the same vein, an advantage of Tversky's calculus is the little weight it accredits to features with a low probability by multiplying the frequency of a feature in X with that in Y. In the case of a main effect of expression type, however, the present approach calculates the size of the weighted shared set as the frequency of shared features in X plus the frequency of shared features in Y: $W_S = {}_XW_S + {}_YW_S$. However, when term is included as a factor, this approach is capable of identifying asymmetry within the shared set: ${}_XW_S$ - ${}_YW_S$. Thus, the function g (Chapter 1) - which should define how to treat the different frequencies in X and Y of a shared feature - simply is the sum in the case that term is not a factor; it is a subtraction when term is a factor.

To ascertain the shared set, two questions should be answered. What are features? When are features equal? If features are listed by the subjects, which informational units should be analyzed: Lines, words, or letters? If two letter strings are equal, are they also semantically equal? The feature 'street' is formally found in 'street car', however, semantically, they do but partly match. On the other hand, the feature 'car' is also found in 'street car', which semantically, makes a better match. Likewise, is 'fisherman' equal to 'fishermen', psychologically speaking?

Similarly, homonyms have different meanings. They are two words, yet, written exactly equal. In other words, formal equality does not necessarily indicate semantic equality as formal inequality does not predicate semantic inequality without more. The problem for any semantic analysis (either heuristic or computational) is that meaning is derived from the form.

Equality of features may be defined in the *strict* sense that formally and semantically, they are identical. By this definition, 'fisherman' and 'fishermen' are not equal, the underlying assumption being that any formal inequality is a semantic inequality. This definition is often implicit in psychological and computational analysis of feature elicitation data.

Equality of features may also be defined in a more moderate sense. The estimate of similarity may neglect formal properties of the features such as affix, suffix, singular or plural. In that case, word-stems are equal to word-derivations. Similarity estimates may be based on 'similitude', on the partial resemblance between elements (Foucault 1988: 57). Using such an *intermediate* definition, formal and semantic equality are allowed to be somewhat divergent. Those features are the same that are 'approximately' equal e.g., when the word-stems are equal. 'Child' then, is identical to 'children'. However, it is also identical to 'child's play'.

Semantic and formal equality are even farther apart, when it concerns letter combinations within words. Features in the form of letter clusters are of interest to scholars of literature who work in the line of Jakobson (1981: 40) and Lotman (1976: 71). They assume that any formal equality should be considered a semantic match. From this point of view, semantic equality is established when a random set of sequential letters is retrieved in any other word, independent of normal meaning connectedness. It has become a common premise of analyzing poetry that meaning is 'created' between words - however remote - when formal equivalences are found. Starting from this definition, 'ear' is not only equal to 'ears' and 'ear ring', but is also equal to 'earl', 'early', 'bears' and 'rearmament'. Reversely, 'ear' is not the only sequence that can be found in 'rearmament'; 'rear', 'arm', 'ma', and 'amen' are also present. This definition of equality is *lenient*, because any formal equality is conceived of as a semantic equality. Using such a definition, the shared set is also composed of equal letter combinations, independent of normal meaning relatedness.

In other words, the computer models of Figure 1.1, Figure 2.1, and Figure 3.1 might distinguish lines, words and letters as features. The *line* is defined as the letter strings and spaces between two 'returns'. The *word* is defined as a letter string between two blanks. A *letter* is every sign that is not a blank. 'Lines' in this sense are not to be understood as necessarily grammatical correct or complete sentences. They may also be genitive constructions, incomplete stock phrases, or chunks of clustered words. Likewise, words may be letter strings not present in the dictionary.

Equality may be either complete or partial. Formally, equality is complete, if each letter of a feature is exactly the same as the letter at the same position in the other feature. Formal equality is partial, if a letter string is equal to a string in another feature, irrespective of surrounding letters or string position. Semantic equality may be found, either by consulting the dictionary (the meaning definitions of the speech community), or by adding a heuristic procedure to this consultation that is not part of the common procedure in the speech community (as in the case of Jakobson's formal-semantic equivalence theory).

Henceforth, the question $x_i = y_j$ *?* in the decision diamond of the *comparison phase* may be rewritten with a combination of 2 formal and 2 semantic equality criteria, calculating overlap for 3 different kinds of features. However, because this would lead to 2*2*3 = 12 analyses for every metaphor model, a combination of feature type and equality criterion was chosen such that the sharpest contrasts occured between measures.

Hence, the first measure (M1) combines the strict definition of equality with lines as definition of features. When two feature sets are compared, this measure tests every word in a line for exact formal equality with the word at the same position in the compared line. In the application of the strict criterion to the computer models, the operator == tests for equality $x_i == y_j$ *?*. This measure respects the way subjects have formulated their associations, and assumes that every formal difference is a semantic difference. M1 finds the smallest number of shared features. The feature set sizes are the number of lines. By definition of the strict criterion, equality is a symmetric relation between the feature sets, unless subjects associate the same feature twice, for example, as both a literal and a figurative feature.

The measure M2 combines a word definition of features with the intermediate criterion of equality. A partial letter string overlap, however, must be meaningful. To decide what meaningfulness is, this measure uses a thesaurus as 'mental lexicon', in which words with their derivations are clustered. If a word in a feature set shows partial overlap with another (for example, larger) word in the other set, then both words must be traced back to the same thesaurus cluster, before they are regarded as a match. Formal equality, then, is based on 'similitude', on the partial resemblance between elements. This looser measure finds more shared features than M1, and implemented in the computer models, it uses the approximate sign $x_i \approx y_j$ *?*. The feature set sizes are numbers of words. Frequency differences of features in the shared set are guaranteed by the fact that one set may contain only the word stem of all the derivations in the other set.

The Jakobsonian criterion is made operative in measure M3. The search is letter oriented, and the demand of meaningfulness is dropped. This measure uses what are termed 'regular expressions' ($/x_i / y_j$ *?*), and looks for letter combinations within word bounderies. The lower limit for a letter combination is the smallest word associated by the subject. Otherwise, every single letter making a match excessively inflates the shared set. Due to its letter orientation, this measure is 'equality-eager'. The number of shared features is expressed as the number of equal letter combinations; the feature set size is the total of letters. Frequency differences within the shared set are the effect of features containing more identical letter combinations in the one feature set than in the other.

To avoid the shared set from growing exorbitantly, a mitigating precaution is taken for M2 and M3. The particles 'de', 'het' and 'een' ('the', 'a', 'an'), the possessive pronouns 'dit', 'dat', 'die', 'deze' ('this', 'that', 'these', 'those'), certain prepositions 'van', 'in', 'om', 'aan', 'bij', 'te' ('of', 'from', 'in', 'on', 'by', 'to'), the logical operators 'en', 'of' ('and', 'or'), certain personal pronouns 'je', 'me', 'we', 'ze' ('you', 'me', 'we', 'they'), the adverb of place 'er' ('there'), and the verb form 'is' ('is') are ignored by the program in counting shared features. These small function words are also discounted for feature set size.

These adjustments enable investigation of the differences in (shared) feature set size among expression types. Moreover, feature set size, or different frequencies of features in the shared set may now explain asymmetry in judgements on similarity.

5.4 **Test of the metaphor models: Feature activation and mapping**

Three between-subject feature elicitation and categorization conditions were administered (*single term*, *expression*, *context*), each containing three expression types (literal, metaphoric, anomalous - within subjects), incorporating an *A*- and a *B-term*. This corresponds to three factors, with three levels each, and two sub-levels within each level.

Subjects were undergraduates of language and literature, Dutch native speakers, aged between 19 and 29 years, who had never previously participated in an experiment. They served as paid volunteers in a free association production task in which as many associations as possible should be generated in response to the terms of the metaphors, literals, and anomalies. Subjects were seated in front of a text processor (486DX2), and reported to have text processing experience. After stimulus presentation, subjects had one minute association time. A tone was the cue to finish the last association, before working on the next stimulus. The inter-trial interval took about 10 to 12 seconds. Trials could consist of a *single term* (one stimulus), an *expression* (both terms forming two stimuli), or an expression in *context* (both terms forming two stimuli). When a trial consisted of an *expression*, the *A-term* was highlighted, prior to the start of the association time. After the tone, the *B-term* was highlighted and the second association minute began. In trials with *context*, subjects read the text at their own speed, indicating by button press that the *A-term* could be highlighted. Further procedures were identical to those for expressions. Associations were automatically stored, every two seconds.

Subjects were asked to type as many associations as possible, with three leads. They were asked to make meaningful associations, and avoid rhyme associations or nonwords. They were also asked to avoid digressing from the meaning of the term by means of 'chain-associations'. However, these rules were not strict, and subjects were free to ignore them. Very personal associations were allowed as long as subjects considered them meaningful features of the stimulus. Subjects were asked to write their associations as catchwords and small sentences, with each new association on a new line. They were unaware of the origin of the stimuli, their connectedness, or the aims of the study. All subjects worked on five training stimuli unconnected to the test set, before starting the actual experiment. Expressions, and their context, were visible during the association period. Trials were mixed in a pseudo-random order. No variants of an expression occurred after each other, and no instances of the same expression type occurred more than twice in a row. Moreover, literal, metaphoric and anomalous variants appeared first, second and last in their relative order of presentation, counterbalanced over expression types. Each subject within a group was alloted a different order of presentation.

In the *single term* condition, twenty subjects received all 85 unique terms from Table 4.1 (Chapter 4) as stimuli. After finishing the association task, subjects decided for each feature whether it was literal or figurative. 'Literal features' were defined as 'belonging to the real world' (as experienced by the subject), and 'figurative features' as 'imaginitive, or symbolic features'. If they felt that a feature was both literal and figurative, it could be copied to the next line, each line receiving a different category letter. Subjects also typed relations behind the features, which could connect a feature to another feature, or to the term. 'Relations' were defined as 'semantic connections between words'. An example from Tourangeau & Sternberg (1982) was used as an illustration. It said that the relation between teachers and kings may be that they both have 'power', the first over his class, the second over his empire. Features already listed as well as newly associated ones were allowed to serve as relations. Subjects were free to write as many relations as possible. However, they did not *have* to. In other words, features not followed by relations remained 'attributive', whereas those receiving a relation were considered 'grounds'. A feature list for a term thus appeared as shown in Table 5.1.[6]

Ten new subjects also worked in the *single term* condition; however, they neither categorized the features as literal or figurative in this condition, nor did they create relations. After finishing *single term*, they entered the *expression* condition, in which they associated on all terms of all the 27 literal, 29 metaphoric and 27 anomalous expressions of Table 4.1. A trial consisted of one expression, containing two stimuli (making 166 stimuli in total). Subsequently, subjects carried out an *expression interpretation* condition, in which they interpreted all expressions by eliciting features for the combination of terms. A trial consisted of one expression, the combination of terms forming one stimulus (making 83 stimuli in total). Thereafter, subjects decided whether the features of each term and of each interpretation of an expression were literal or figurative. Moreover, they constructed relations between features or between stimuli and features.

Table 5.1: Results for the stimulus **heart** by subject 5 in the *single term* condition, illustrating the creation of a small feature set, after categorizing the features for a term as literal (l) and/or figurative (f), and making relations (r) between them. Feature elicitation was performed before l- and f-categorization and relation creation. English translation from the Dutch.

heart

red l
body l
love f
blood l
bloodred l
nursery l sweet r
sweet f
valentine l
spear l death r
sword l death r
fight l death r
war l death r
cow's heart l
food l
dog l food r
passionate f
bosom friend f
fruit hearts l food r

Another ten subjects also received the *single term* condition, whereupon the same procedure was followed as described in the previous paragraph, except that in *context*, the original poem was added to the expressions. A trial consisted of an expression in context, containing the two terms as stimuli. Similarly, in the *context interpretation* condition, the combination of the terms in the expressions was interpreted in context. One trial was composed of an expression in context, the combination of both terms forming one stimulus. As described above, literal, figurative and relation features were attached, thereafter.

A *single term* condition preceded *expression* and *context*, because in the latter two conditions certain terms were repeated in the stimulus set (e.g., 'death', 'heart'), whereas others were not (e.g., 'wall'). Moreover, all variants of an expression repeated the *A-term*. This could lead to a learning advantage for repeated terms over unrepeated terms. More features could be activated, which increases the probability to form overlap. Presenting *single term* in advance might spread repetition effects more equally over all terms, so that if not eliminated, its effect was neutralized at least. Controls on the repetition effect will be part of the discussion.

Before entering the analysis, the raw data were run through a spelling checker; misspellings were repaired, and alternative spellings were made uniform. The data were analyzed according to the metaphor models of Section 1.3, 2.3 and 3.3, using three measures of feature overlap: M1, M2, and M3.

Deciding among the measures

As argued above, it is unclear beforehand what formally should be considered a feature in the association data (the lines, the words, the letter clusters?), and what the criterion for feature equality should be (strict, intermediate, lenient?). Which measure (M1, M2, M3) to use, then? In the next paragraphs, a statistical test is exploited to decide among the measures.

The distinctive power of each measure was investigated, to opt for one and to omit all others. Since the measures M1, M2, M3 are concerned with six within-variables each (two terms and three expression types), 18 within-variables were treated as repeated measures in a MANOVA. Discriminant functions for M1, M2, and M3 decided which measure differentiated best among conditions, expression types and terms. The measure with the highest partial regression weight (or coefficient) was supposed to make the subtlest distinctions among effects. When ties occurred between two or more weights, the highest correlation was decisive.

Table 5.2 (see Appendix) exhibits the effects on the feature sets relevant to the metaphor models, along with the discriminant functions and correlations. For each effect, only the first discriminant functions were considered, because potential functions for the residuals are equally or less important. Bonferroni correction was disregarded, and marginally significant effects ($p \gtrsim .05$) were analyzed for regression weights, to maximize the information about the distinctive power of the measures. Measures with the highest distinctive power for an effect are labeled with an asterisk.

Reference to Table 5.2 reveals that *M1 = 9, *M2 = 20 and *M3 = 7. M2 most often received the highest partial regression weights (20 times). Thus, the intermediate definition of equality combined with words as features (M2) distinguished most factors within effects, whereas M3 distinguished least. Therefore, **the analysis of the feature elicitation data will be performed only with M2.**

Results of feature elicitation, l-f categorization and r-creation

Feature sets and shared sets were not accumulated across subjects, but analyzed within subjects. The data shown in Figure 5.0 up to 5.5 were subjected to MANOVA (Appendix Table 5.3). For the main effects of condition and term, unique sums of squares were computed, whereas all other effects were determined by Pillai's Trace. To accommodate Clark's (1973) critique on language statistics,[7] all effects significant with subjects as random factor were also tested against a quasi-F with stimuli as the random factor. All significant effects were followed by parameter estimates.

As proposed in Hoorn (in press), F-values were tested against $\alpha = .05$, whereas t-tests according to Bonferroni handled an $\alpha = .05$ for the comparison of two means (main effect of term), an $\alpha / 3 \approx .0166$ for clusters of three related effects (main effect of condition, main effect of expression type, condition by term, term by expression type) and an $\alpha / 9 \approx .0055$ for clusters of nine related effects (condition by expression type, condition by term by expression type). Henceforth, * = $p < .05$, ** = $p < .0166$, *** = $p < .0055$, unless indicated otherwise.

Results of weighted feature set size

Figure 5.0: Grand mean weighted feature set size W_X (solid) and W_Y (dashed). Features are words. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

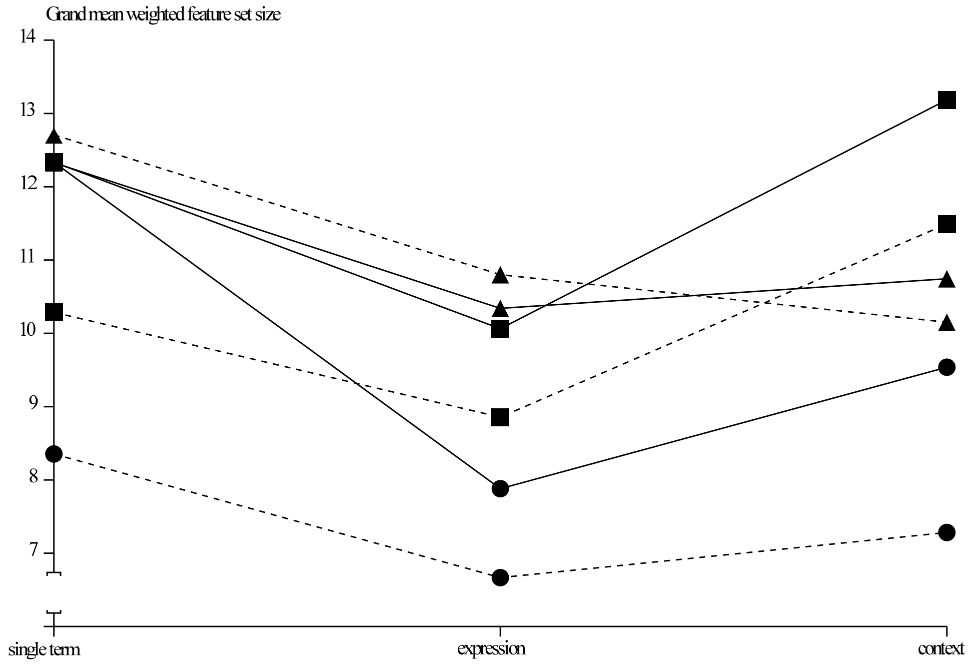


Figure 5.0 presents the total number of weighted features averaged over subjects (grand means) for terms, expression types and conditions. The fol-

lowing observations were derived from the significant second-order interaction among these three factors (Appendix Table 5.3: ❶$_a$).

(I) *Expression* tended to elicit fewer features than *single term* and *context*, although only the difference with *single term* was statistically reliable. (II) Metaphors benefited little from context. (III) *A-terms* tended to elicit more features than *B-terms*, with the exception of metaphors in *single term* and *expression*. (IV) Anomalous *B-terms* elicited the least features.

Figure 5.1: Grand mean weighted literal feature set size W_{Xl} (solid) and W_{Yl} (dashed). Features are words. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

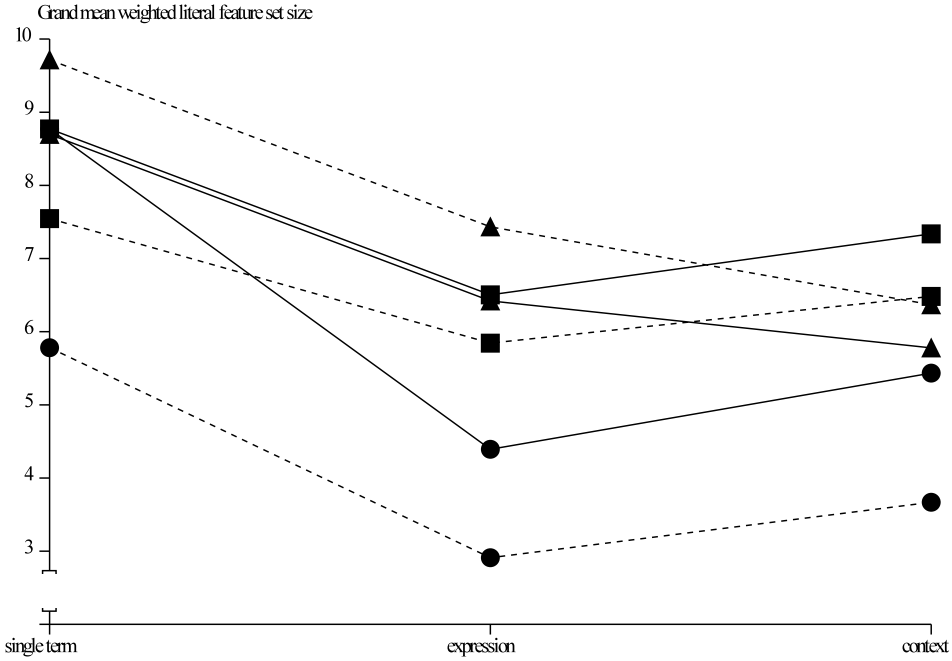


Figure 5.1 exhibits the grand means for those features classified as literal (W_{Xl} , W_{Yl}). The following observations were derived from the significant effects in Table 5.3 (Appendix).

(I) *Single term* activated more literal features than *expression* and *context*. However, only the difference with *expression* was reliable (main effect of condition: ❷$_a$). (II) Anomalies yielded the least number of literal features, whereas literal and metaphoric expressions were about equal (interaction of term by expression type: ❸$_a$). (III) There were more literal features for *A-* than for *B-terms* in literal and anomalous expressions, whereas the difference reversed in metaphors (interaction of term by expression type: ❸$_a$).

Figure 5.2: Grand mean weighted figurative feature set size W_{Xf} (solid) and W_{Yf} (dashed). Features are words. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

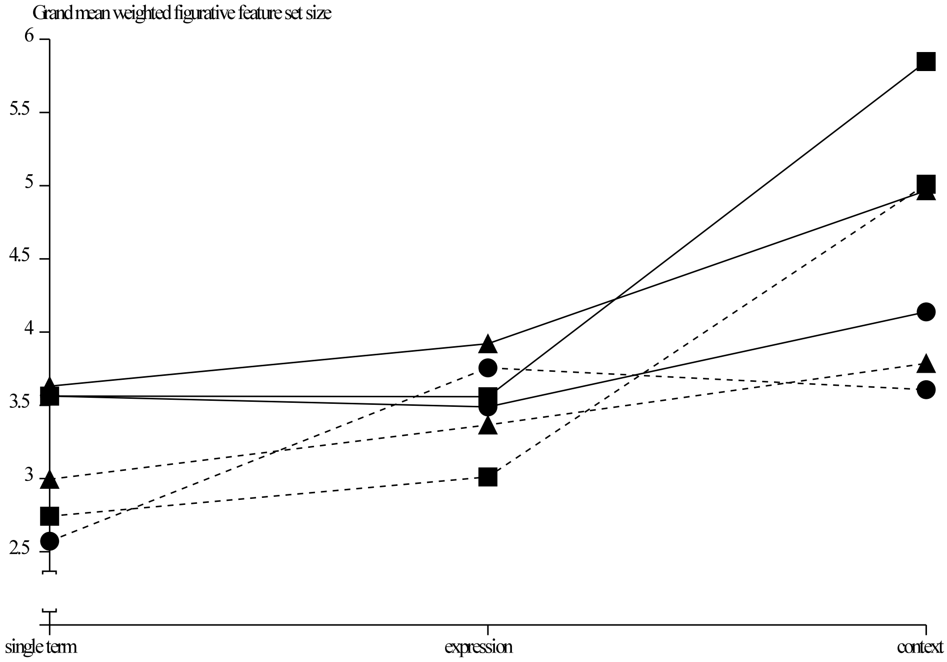
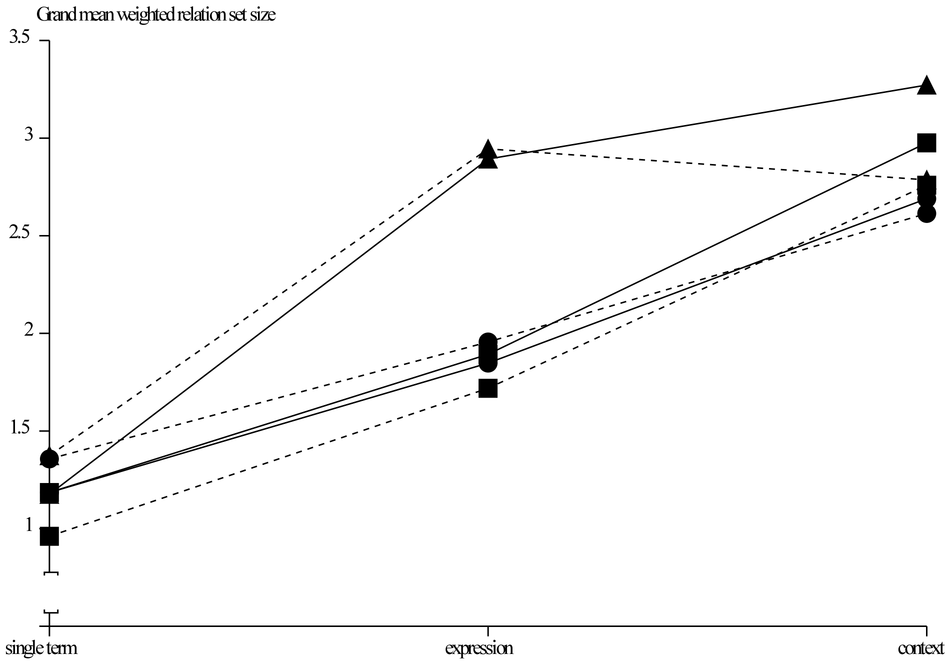


Figure 5.2 displays the grand means for those features classified as figurative (W_{Xf} , W_{Yf}). According to Table 5.3 (Appendix), only the main effect of term was significant (❹$_a$), indicating that:

(I) *A-terms* raised more figurative features than *B-terms*. Although it seems that more figurative features are activated when more context is added, the main effects of condition proved unreliable.

Figure 5.3: Grand mean weighted relation set size W_{Xr} (solid) and W_{Yr} (dashed). Features are words. ■ = literal expressions, ▲ = metaphors, ● = anomalies.



The application of relations was optional, so that certain features were followed by relations (cf. Table 5.1), whereas others were not. Those features *x* and *y* not ensued by relations *r* were considered 'genuine' attributes *a*; those connected with relations were considered grounds *g* (note that the notion 'ground' only means that the subject connected this feature with a relation. This is different from the conventional use of 'ground'). Thus, the following weighted set sizes were analyzed: (W_{Xr} , W_{Yr}), (W_{Xg} , W_{Yg}) and (W_{Xa} , W_{Ya}).

In Figure 5.3, the grand mean weighted relation set size W_{Xr} and W_{Yr} are plotted. Only one observation was siginificant, namely the main effect of condition (❺$_a$), showing that:

(I) The number of created relations increased with the increase of context. *Context* evoked more relations than *single term*. Despite the apparent sensitivity of metaphors for creating relations in *expression* and *context*, no effects involved with expression type were statistically reliable.

Figure 5.4: Grand mean weighted ground set size W_{Xg} (solid) and W_{Yg} (dashed). Features are words. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

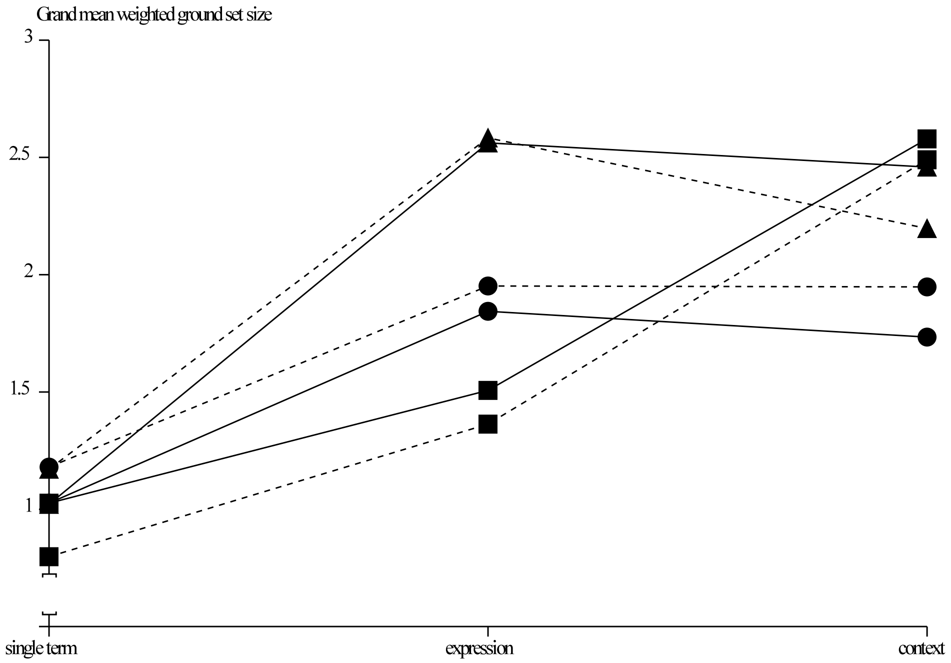


Figure 5.4 shows the grand mean weighted ground set sizes W_{Xg} and W_{Yg} . As in the case of the relations, only the main effect of condition was significant (❻$_a$), showing that:

(I) The number of grounds increased with the increase of context. In *context*, more grounds were found than in *single term*. Again, a special sensitivity of metaphors was not materialized by a significant effect of expression type.

Figure 5.5: Grand mean weighted attributive feature set size W_{Xa} (solid) and W_{Ya} (dashed). Features are words. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

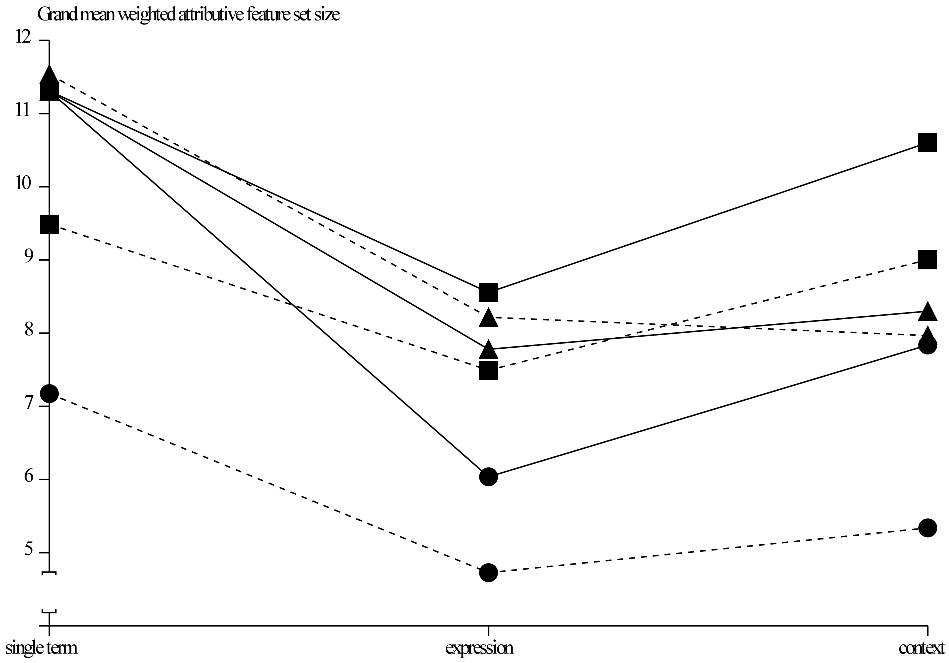


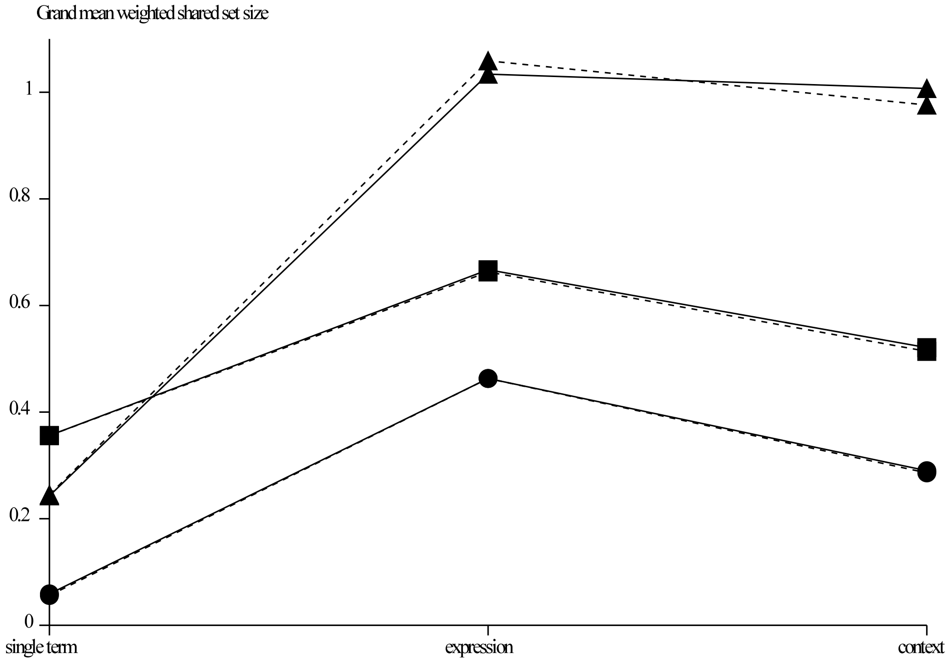
Figure 5.5 shows the grand mean weighted attributive feature set sizes W_{Xa} and W_{Ya} . The significant second-order interaction (❼$_a$) followed exactly the same pattern as in Figure 5.0.

(I) In *expression*, fewer features were elicited than in *single term* or *context*. However, only the difference with *single term* was significant. (II) Metaphors benefited little from context. (III) *A-terms* evoked more features than *B-terms*, with the exception of metaphors in *single term* and *expression*. (IV) Anomalous *B-terms* raised least features.

Results of weighted shared set size

Analogous to Table 5.3, Table 5.4 (Appendix) contains MANOVAs for the effects shown in Figure 5.6 up to 5.13.

Figure 5.6: Grand mean weighted shared set size W_S, using M2. $_XW_S$ (solid) is the number of weighted features that the *A-term* shared with the *B-term*. $_YW_S$ (dashed) is the number of weighted features that the *B-term* shared with the *A-term*. ■ = literal expressions, ▲ = metaphors, ● = anomalies.



In Figure 5.6, measure M2 calculated the weighted shared set as the number of weighted features that the *A-term* shared with the *B-term* ($_XW_S$), and the number of weighted features that the *B-term* shared with the *A-term* ($_YW_S$).

(I) *Single term* shared fewer features than *expression* and *context* (main effects of condition: ❶$_b$). (II) Literal expressions and metaphors shared more features than anomalies (main effects of expression type: ❷$_b$). Moreover, Figure 5.6 suggests that in *single term*, literals shared more features than metaphors, which shared more than anomalies. By contrast, metaphors yielded more shared features than literals, which shared more features than anomalies in *expression* and *context*. However, these interactions of condition by expression type remained unreliable. (III) Feature frequency differences in the shared set were found in *expression* (*A-term* < *B-term*) in contrast with *context* (*A-term* > *B-term*) (interaction of condition by term: ❸$_b$).

Figure 5.7: Grand mean weighted literal shared set size (W_{SIl}), using M2. ${}_XW_{Sl}$ (solid) is the number of weighted literal features that the *A-term* shared with the *B-term*. ${}_YW_{Sl}$ (dashed) is the number of weighted literal features that the *B-term* shared with the *A-term*. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

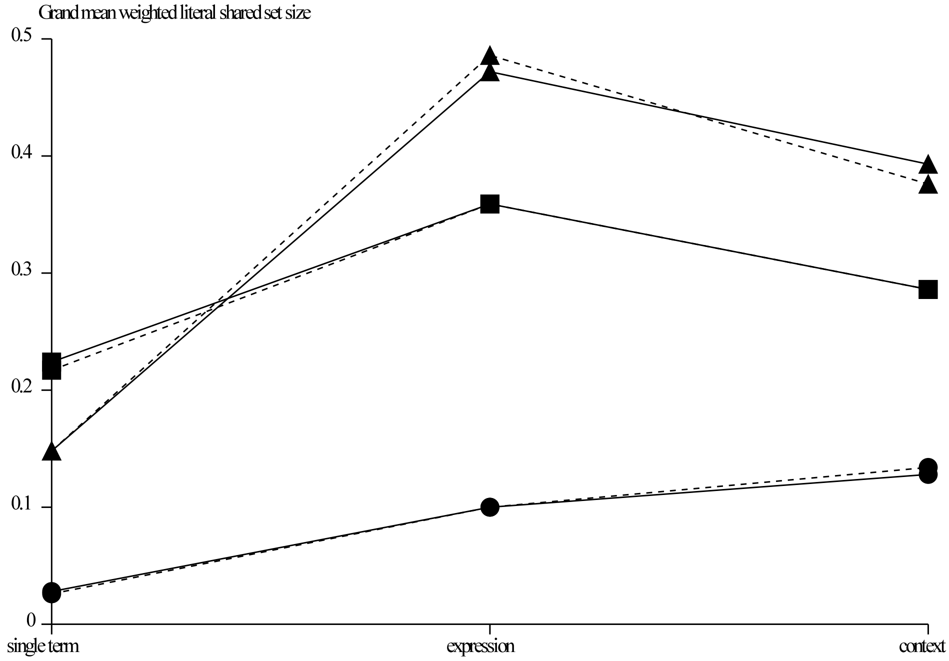
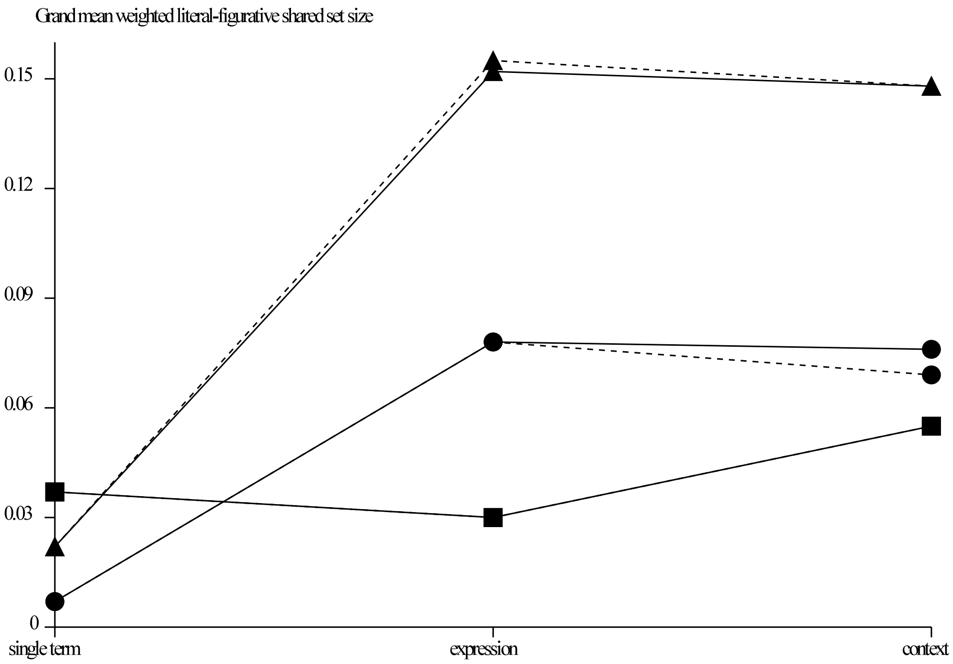


Figure 5.7 provides the grand mean weighted shared set sizes for literal features (W_{SIl}). Observe the following effects, validated by Table 5.4 (Appendix).

(I) In *single term*, smaller literal shared sets were found than in *expression* or *context* (main effects of condition: ❹$_b$). (II) Anomalies rendered the smallest literal shared sets compared with literal expressions and metaphors, which were about equal (main effects of expression type: ❺$_b$). The apparent increase of the literal shared set for metaphors in *expression* and *context* was not manifested in a significant interaction between condition and expression type.

Figure 5.8: Grand mean weighted literal-figurative shared set size (W_{Slf}), using M2. $_{X}W_{Sl}$ (solid) is the number of weighted literal features that the *A-term* shared with the figurative features of the *B-term*. $_{Y}W_{Sf}$ (dashed) is the number of weighted figurative features that the *B-term* shared with the literal *A-term* features. ■ = literal expressions, ▲ = metaphors, ● = anomalies.



For the results in Figure 5.8, the grand means were calculated of the weighted shared features (W_{Slf}) categorized as literal for the *A-term* and figurative for the *B-term*.

(I) *Single term* had the smallest shared sets of literal *A-term* and figurative *B-term* features compared with *expression* or *context* (main effects of condition: ❻$_{b}$). (II) Metaphors elicited the largest sets of shared features that were literal for *A-terms* and figurative for *B-terms* (main effects of expression type: ❼$_{b}$). (III) In *single term*, the literal-figurative shared set was larger for literal expressions than for metaphors. Reversely, it was larger for metaphors than for literal expressions in *expression* (interaction of condition by expression type: ❽$_{b}$).

Figure 5.9: Grand mean weighted figurative-literal shared set size (W_{Sfl}), using M2. $_XW_{Sf}$ (solid) is the number of weighted figurative features that the *A-term* shared with the literal features of the *B-term*. $_YW_{Sl}$ (dashed) is the number of weighted literal features that the *B-term* shared with the figurative *A-term* features. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

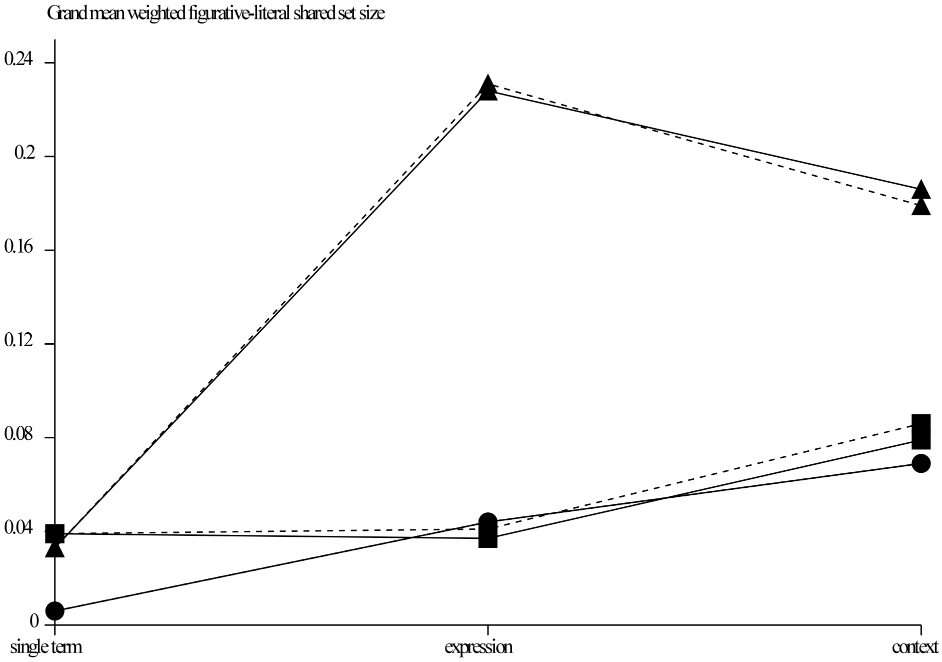


Figure 5.9 presents the grand means of the weighted shared features (W_{Sfl}) categorized as figurative for the *A-term* and literal for the *B-term*. Observe that the results are identical to those for W_{Slf}, except for point (III).

(I) The smallest shared sets of figurative *A-term* and literal *B-term* features were found in *single term* compared with *expression* or *context* (main effects of condition: ❾$_b$). (II) Metaphors invoked the largest sets of shared features that were figurative for *A-terms* and literal for *B-terms* (main effects of expression type: ❿$_b$). (III) In *single term*, the figurative-literal shared set was a little larger for literal expressions than for metaphors. Invertedly, it was larger for metaphors than for literal expressions in *expression*. Moreover, it was larger for metaphors than for anomalies in *single term*, which increased in *expression* (interaction of condition by expression type: ❶$_c$).

Figure 5.10: Grand mean weighted figurative shared set size (W_{Sff}), using M2. ${}_XW_{Sf}$ (solid) is the number of weighted figurative features that the *A-term* shared with the *B-term*. ${}_YW_{Sf}$ (dashed) is the number of weighted figurative features that the *B-term* shared with the *A-term*. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

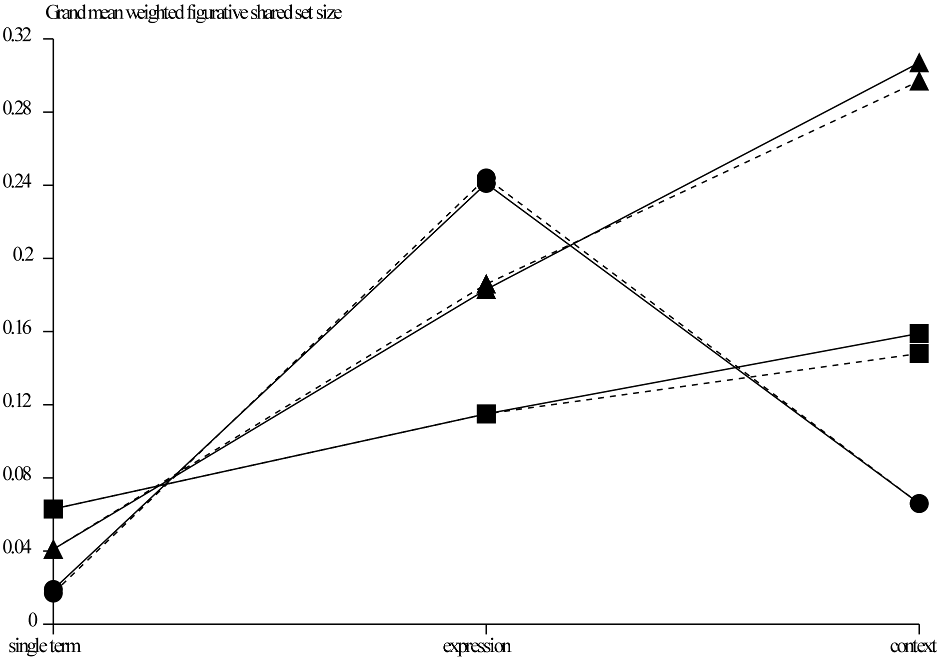


Figure 5.10 offers the grand mean weighted shared set sizes of figurative features (W_{Sff}). The following observation was justified by the significant main effect of condition (❷$_c$).

(I) *Expression* and *context* yielded larger figurative shared sets than *single term*.

Most conspicuously, anomalies benefited from *expression*. However, no effect concerning expression type was significant. Accordingly, the apparent rise of shared figurative features for metaphors in *context* was insignificant.

As pointed out above, the features x and y could be attributes a or grounds g, depending on the presence of a relation r or not. If this differentiation is neglected, relations are just shared with x and y. In other words, the shared set can be calculated for the following sixteen combinations of set types:

$$
\begin{array}{llll}
(X \cap Y), & (X \cap Y_r), & (X \cap Y_g), & (X \cap Y_a) \\
(X_r \cap Y), & (X_r \cap Y_r), & (X_r \cap Y_g), & (X_r \cap Y_a) \\
(X_g \cap Y), & (X_g \cap Y_r), & (X_g \cap Y_g), & (X_g \cap Y_a) \\
(X_a \cap Y), & (X_a \cap Y_r), & (X_a \cap Y_g), & (X_a \cap Y_a)
\end{array}
$$

Since $X \cap Y$ was already analyzed for Figure 5.0, it was replaced by the intersection of complete relation+ground (*rg*) combinations: ($X_{rg} \cap Y_{rg}$). Examples of relation+ground combinations are found in Table 5.1. 'Sweet' is a relation to the ground 'nursery', 'death' is a relation to the ground 'spear'.

Preliminary MANOVAs showed that the weighted shared set (W_{Sgr}) between grounds of the *A-term* (${}_XW_{Sg}$) and relations of the *B-term* (${}_YW_{Sr}$) rendered equal or better results than W_{Srr} and W_{Sxr} (the original interaction model variables). The other kinds of weighted shared sets obtained fewer effects than W_{Srr} , W_{Sxr} or W_{Sgr} and were dismissed from further analysis.

Figure 5.11: Grand mean weighted relation shared set size (W_{Srr}), using M2. ${}_XW_{Sr}$ (solid) is the number of weighted relations that the *A-term* shared with the *B-term*. ${}_YW_{Sr}$ (dashed) is the number of weighted relations that the *B-term* shared with the *A-term*. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

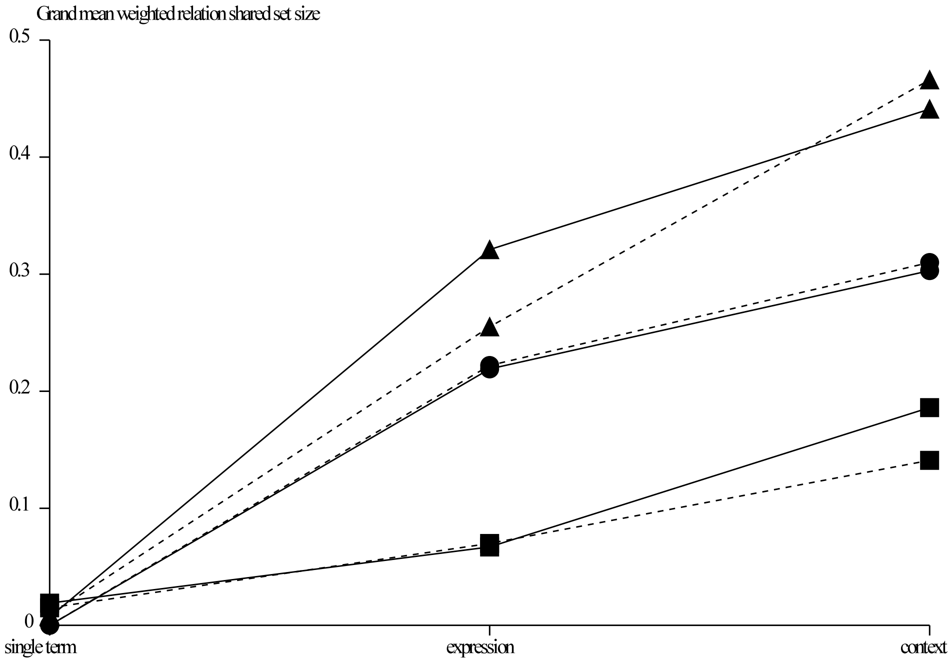


Figure 5.11 exhibits the grand mean weighted set sizes of shared relations (W_{Srr}). The following significant main effect of condition (❸$_c$) was the only effect that occurred:

(I) The number of shared relations increased as a function of additional context, because *expression* and *context* evoked more shared relations than *single term*.

Since this was the only significant effect, the increase of shared relations for metaphors - with a particular high contribution of the *B-term* - in *context* is unreliable.

Figure 5.12: Grand mean weighted feature-relation shared set size (W_{Sxr}), using M2. ${}_{X}W_{Sx}$ (solid) is the number of weighted features that the *A-term* shared with the *B-term* relations. ${}_{Y}W_{Sr}$ (dashed) is the number of weighted relations that the *B-term* shared with the *A-term* features. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

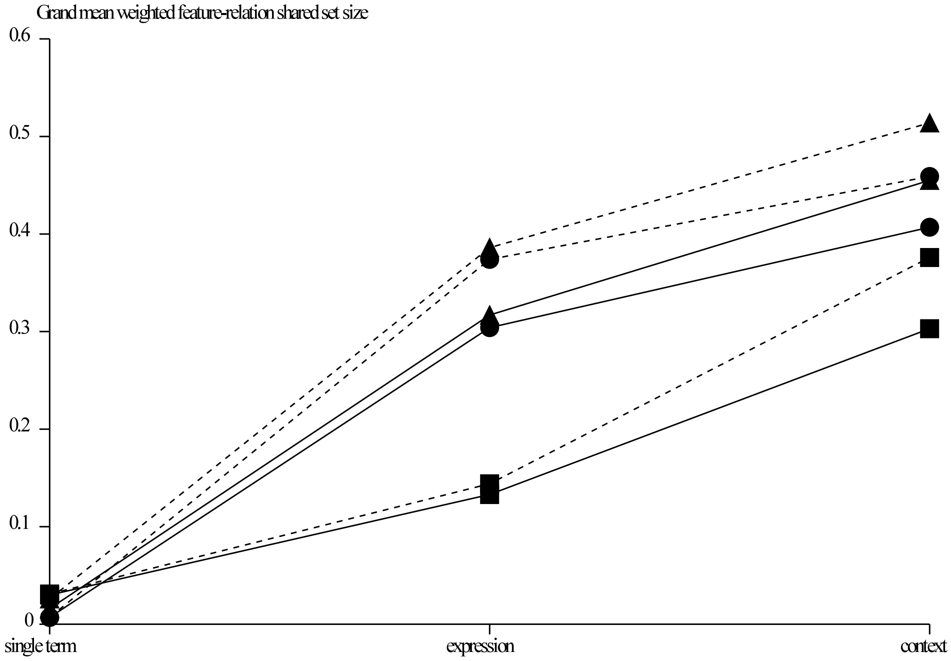


Figure 5.12 shows the grand mean weighted shared set of the *A-term* features - not differentiated in attributes and grounds - and *B-term* relations (W_{Sxr}). The observations following next were substantiated by the significant interactions of condition by term (❹$_c$).

(I) The shared set between *A-term* features and *B-term* relations grew when contexts were extended. *Single term* yielded less overlap than *expression* or *context*. (II) The frequencies of shared *B-term* relations were higher than the frequencies of the corresponding *A-term* features. Thus, *B-term* relations contributed more to the shared set. This effect was stronger in *expression* and *context* than in *single term*.

Although metaphoric *B-term* relations in *context* augmented the shared set more than the other term types in whichever condition, no effect containing expression type was reliable.

Figure 5.13: Grand mean weighted ground-relation shared set size (W_{Sgr}), using M2. ${}_{X}W_{Sg}$ (solid) is the number of weighted grounds that the *A-term* shared with the *B-term* relations. ${}_{Y}W_{Sr}$ (dashed) is the number of weighted relations that the *B-term* shared with the *A-term* grounds. ■ = literal expressions, ▲ = metaphors, ● = anomalies.

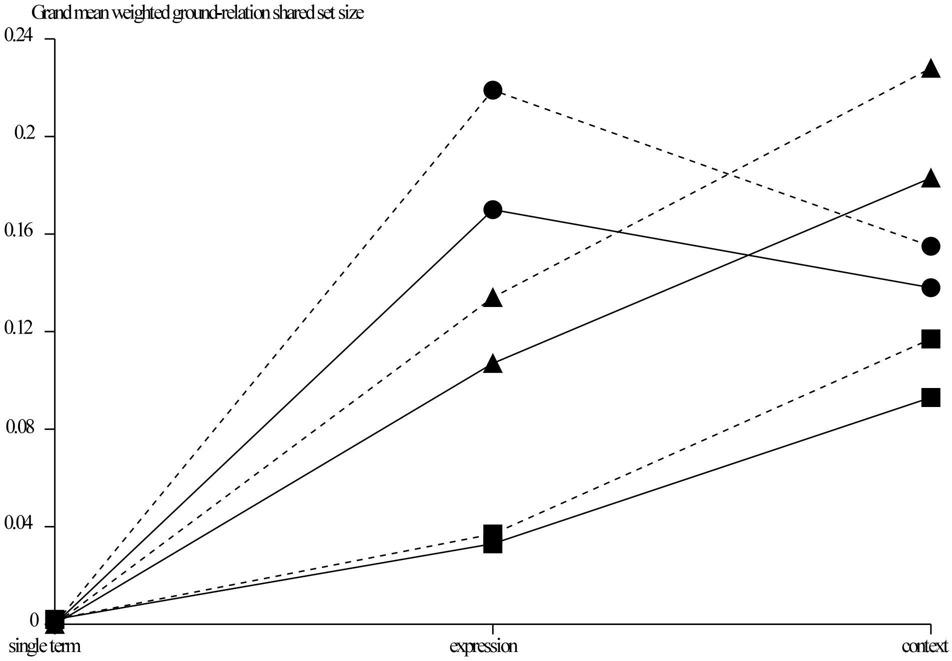


Figure 5.13 shows the grand mean weighted shared set between grounds of the *A-term* and *B-term* relations (W_{Sgr}). The same observations as for W_{Sxr} (Figure 5.12) are valid, according to the interactions of condition by term (❺$_c$).

(I) The set of *A-term* grounds shared with *B-term* relations increased with the extension of context. *Expression* and *context* raised larger shared sets than *single term*. (II) Shared *B-term* relations had higher frequencies than the corresponding *A-term* grounds, stronger in *expression* and *context* than in *single term*.

Despite the apparent higher number of metaphoric *B-term* relations in *context* and the drop for anomalies in *context* compared with *expression*, effects involving expression type were insignificant.

Discussion of feature set size and shared set size

The major issue of this discussion is to confront the predictions of the metaphor models with the above results. The comparison model assumes fixed feature sets for the individual terms, irrespective of condition. Moreover, subjects decide whether an expression is literal, metaphoric or anomalous on the basis of the shared set size. Literals should have larger shared sets than metaphors, which should have larger shared sets than anomalies. To what extent do these predictions agree with the results?

First, feature set size was not fixed, but depended rather on condition and expression type (Figure 5.0). *Single term* evoked significantly more features than *expression* and *context*; terms yielded fewer features when they were part of anomalies than when part of literals and metaphors, and the sign of the ratios between W_X and W_Y reversed relative to these factors. In principle, this is at odds with the comparison model. However, it could be argued that although the terms have a fixed feature set, not all features are actually produced. Although all features might have been activated, not all were written down by the subjects.

Relatively good news came from the weighted shared set W_S (Figure 5.6). The comparison model expects main effects of expression type on the shared set, without interactions with condition. Indeed, no such interaction was statistically reliable. However, shared sets should be fixed, which was not the case. The main effect of condition showed that *single term* shared fewer features than *expression* and *context*. It is essential for the comparison model that the shared set for literals is larger than for metaphors, which in its turn is larger than for anomalies. However, the main effects of expression type demonstrated that differences in shared set size between literals and metaphors were insignificant, and that only anomalies had smaller shared sets.

In sum, **feature sets and shared sets were not fixed. *Single term* yielded the most features, but the smallest shared sets. *Expression* yielded the smallest feature sets; however, also the largest shared sets. *Context* yielded both large feature sets and large shared sets. Although these results are incongruous with the comparison model, they are not decisive and may be due to production failures. What is essential is that literal expressions did not have larger shared sets than metaphors, although both had larger shared sets than anomalies.**

The anomaly model expects that the weighted set sizes of literal (W_{Xl} , W_{Yl}) and figurative (W_{Xf} , W_{Xf}) features are fixed and that literal expressions do not activate figurative features. Furthermore, literal expressions share most literal features and no figurative ones. Metaphors share most figurative features, whereas anomalies none of either type. Strictly speaking, the shared sets should also be fixed.

Thus, the same predictions for feature set size are valid as for the comparison model with the same constraint that feature activation does not have to be identical to reporting features. Main effects of expression type and term (or their interactions) are permitted, but they are not essential. Main effects of condition or interactions with condition are evidence against fixed sets.

Figure 5.1 broadly followed the results of Figure 5.0, so that indeed the premise on fixed sets should be revised. Again, *single term* showed more literal features than *expression*. Furthermore, significant interactions of term by expression type demonstrated that the sign of the ratios between W_{Xl} and W_{Yl} reversed as a function of expression type. Other interactions were statistically unreliable, in line with the anomaly model.

The results for figurative feature set size (Figure 5.2) show few effects, considering that only the main effect of term was significant. *A-terms* activated more figurative features than *B-terms*, which is immaterial for the anomaly model. All interactions on figurative set size were insignificant, which is consistent with the anomaly model.

More importantly, the anomaly model demands that literal expressions have the largest weighted literal shared sets (W_{Sll}), whereas metaphors and anomalies should have smaller ones. Metaphors should have the largest figurative shared sets (W_{Sff}), whereas literal expressions should have none, and anomalies perhaps small ones. These patterns should be demonstrated in the main effects of expression type. Since the shared set sizes are supposedly fixed, no interactions with the condition factor should be significant. Main effects of term or interactions of term by expression type are irrelevant for the anomaly model.

The grand mean literal shared sets in Figure 5.7 followed the pattern of Figure 5.6. Thus, literal set sizes were not fixed. The main effects of condition on W_{Sll} display that *single term* evoked less literal overlap than other conditions. The main effects of expression type show that literal expressions and metaphors yielded about equal literal overlap, which discredits the anomaly model. In favor are the fact that anomalies yielded the smallest literal overlap and that none of the interactions were significant.

Evidence argued against anomaly theory in the case of the weighted figurative shared set size (W_{Sff}). The crucial main effect of expression type was deficient, since the figurative shared set was not larger for metaphors than for literals or anomalies.

The good news came from two new variables. Besides the purely figurative shared set W_{Sff} , Figure 5.8 and Figure 5.9 exhibit the results of W_{Slf} and W_{Sfl} . For these mixed kinds of overlap, metaphors did yield the largest shared sets, as expected by the anomaly model. Only once was an unpredicted pattern found: W_{Slf} was larger for literal expressions than metaphors in *single term* (interaction of condition by expression type). Nevertheless, since the mixed shared sets contained a figurative component, they may be proper substitutes for purely figurative shared sets.

Recapitulating, **literal and figurative feature sets were not fixed and neither were the shared sets. Literal set sizes were about equally large for literal and metaphoric expressions, and smaller for anomalies. Figurative set sizes and purely figurative overlap did not differentiate the expression types, which makes them disposable. Purely literal overlap could distinguish literals and metaphors only from anomalies, which had smaller sets. Shared sets composed of literal and figurative features, however, could distinguish metaphors from literals and anomalies, because the latter two expression types had smaller shared sets.**

According to the interaction model, a small number of relations is created for literal expressions, whereas considerably more are created for metaphors and anomalies. The creation of relations is mainly a quality of the *B-term* in response to the context. In *single term*, no relations are created, because the expression type cannot be identified. In *expression* and *context*, however, it can, so that in the case of metaphors (and anomalies), many relations are created. Thus, feature sets are not fixed. In effect, all factors should interact, so that fewer relations are created in *single term* than in *expression* and *context*. With respect to these latter two conditions, metaphors and anomalies should stimulate more relations than literal expressions, while the *B-term* ought to dominate the *A-term*. In *single term*, no such differences should be found.

Hence, the second-order interaction of condition by term by expression type should be significant for relation set size. This should be the case, particularly when contrasting the *A-* and *B-term* in *single term* versus *expression* or *context*, and when contrasting literal expressions versus metaphors or anomalies (parameters 1, 2, 4, 5, Table 5.3, Appendix). Second-order interactions contrasting *expression* versus *context* or metaphors versus anomalies may remain insignificant (parameters 3, 6, 7, 8, 9, Table 5.3), because the interaction model does not explicitly predict differences for these contrasts.

When the parameters 3 and 6 are insignificant, the interaction of term by expression type (parameters 1 and 2) should be present. When 7 and 8 are unreliable, condition by term should be operative (parameters 1 and 2). Parameter 9 is not decisive for the interaction model, nor the underlying interactions of term by expression type (parameter 3) and condition by term (parameter 3).

Figure 5.3 shows that for metaphors, most relations were created in *expression* and *context*. However, the main effects of expression type and the interactions with condition were insignificant. Only the main effect of condition was reliable. In *context*, more relations were created than in *single term*, while *expression* was intermediate. This underscores the prediction of the interaction model that relation creation increases when contexts are extended. On the other hand, *B-terms* did not stimulate more relations than *A-*

terms, nor did metaphors in comparison with literals and anomalies, which straightforwardly argues against the interaction model.

Instead, differences among expression types and between terms were due to features other than relations. One variable that could distinguish among expression types was (W_X , W_{Yr}), which compares the weighted set size of the features in X with the weighted relation set size in Y. Since this is a variable that contains relations in Y_r , it may seem that relations were important for the expression type differences. Yet, this was not the case. The main effect of expression type and the interaction of term by expression type on (W_X , W_{Yr}) (Table 5.3) showed that literals and metaphors had larger sets than anomalies. However, another variable (W_{Xr} , W_{Yr}) pointed out that W_{Yr} was not significantly different from W_{Xr} (no effects of term). In other words, W_{Xr} and W_{Yr} were identical and their combination yielded no effects for expression type. Thus, the expression type differences obtained by (W_X , W_{Yr}) were not due to the relations in Y_r , but rather to the features in X.

Another variable that established expression type differences was (W_{Xa} , W_{Ya}), which also does not contain relations. These sets of attributive features obtained second-order interactions (Table 5.3) similar to (W_X , W_Y). Since (W_X , W_Y) is a variable defined by the comparison model, the interaction model prerequisite of completely interactive factors was fulfilled by the comparison-like variable (W_{Xa} , W_{Ya}).

Considering the effects on the shared sets, the interaction model expects that literal and anomalous expressions have no shared relations, whereas metaphors have many. Differences between *A*- and *B-term* are not required per se, but *expression* and *context* should enlarge relation overlap compared with *single term*. Therefore, second-order interactions or interactions of term by expression type should show that in *single term* (L = M = A), whereas in *expression* and *context* (L < M > A). In addition, *context* should elicit equal or higher numbers of shared relations than *expression*.

Figure 5.11 suggests that metaphors share most relations, which increases in *expression* and *context*. However, Table 5.4 (Appendix) indicates that for W_{Srr} , W_{Sxr} and W_{Sgr} , the interactions of condition by expression type were insignificant, and that the main effects of expression type on W_{Srr} were just not below chance ($F_{2,36}$ = 3.13, p = .055). In other words, no effects including expression type were significant for shared relations.[8]

Nonetheless, the Figures 5.11 up to 5.13 show that the number of shared relations increased when context was added. Solid support for this prediction of the interaction model was found in the main effects of condition on W_{Srr} , W_{Sxr} and W_{Sgr} . Relations intersected more in *expression* and *context* than in *single term*. Moreover, interactions of condition by term on W_{Sxr} and W_{Sgr} showed that in *single term* (${}_XW_{Sx\ or\ g} \leq {}_YW_{Sr}$); whereas for *expression* and *context* (${}_XW_{Sx\ or\ g} << {}_YW_{Sr}$). In other words, many relations of the *B-term* were linked with fewer features (or grounds) of the *A-term*, relative to condition, but not to expression type.

To summarize, **there were certain tendencies that metaphors activated and joined relations more than other expression types as a function of context extension. However, except for the main effects of condition, all these effects were statistically insignificant. In conclusion, metaphoric *B-terms* did not activate more relations than other term types, irrespective of condition. On the contrary, expression type differences were established by features that were not relations**.

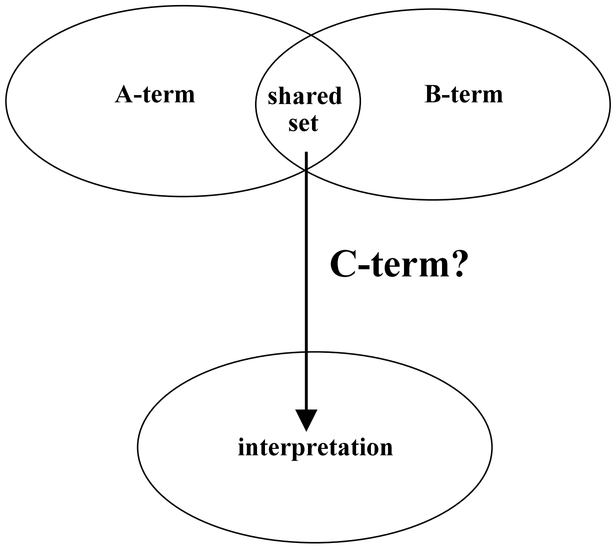
It is obvious that none of the predictions of the models were clearly confirmed. Some were found, although never without contradictory evidence. Regarding the comparison model, literal expressions did not have larger shared sets than metaphors, although both had larger shared sets than anomalies. Concerning the anomaly model, figurative set sizes and purely figurative shared sets did not show expression type differences. Literal shared sets could distinguish literal expressions and metaphors only from anomalies. Shared sets that mixed literal and figurative features distinguished metaphors from literal expressions and anomalies. With reference to the interaction model, no more relations were activated or connected for metaphoric *B-terms* than for other term types, not even in *context*.

Is the shared set identical to the interpretation?

In comparison with literal and anomalous expressions, it seems that metaphors are interpreted by mixes of literal and figurative features. However, is there more to interpreting a metaphor than finding the shared set? The comparison and anomaly theory maintain that the *C-term* - the solution to the metaphor - is (in) the shared set. The interaction model claims that the interpretation of metaphors is at least *constituted* on shared relations, whereas for the other expression types, it is not. Thus, the features in the shared set should be identical to those elicited during an *interpretation* of an expression. Henceforth, the feature set produced during an interpretation will be called Z, and according to the comparison and anomaly model, $X \cap Y = Z$ (cf. Figure 5.14).

As described earlier, subjects in the *expression* and *context* conditions, also generated features for the complete expressions. They were asked not to associate separately on the *A*- and *B-term*, but to *interpret* the expressions, in the same way as for their poetry classes. After feature elicitation, subjects categorized the features in Z as literal and/or figurative, and made relations.

Figure 5.14: The retrieval of shared features in the interpretation. Is the shared set - either literal, figurative, relational or undifferentiated - identical to the interpretation of an expression?



In line with the comparison and anomaly theory, the shared set should be the same as *Z*. Although interaction theory does not claim fixed sets, at least one relation in the shared set should be retrieved in *Z*. Thus, The *C-term* should be one or more features in the shared set *T*, so that $T = X \cap Y \cap Z$. The size of this shared set is $\#T$. The weight *w* of feature *u* in *T* is a function *h* of the weight of *u* in *X*, *Y*, and *Z*: $w_T(u) = h(w_X(u)\ ,\ w_Y(u)\ ,\ w_Z(u))$. The weighted set size (*W*) of all shared features *u* in *T* is the sum of all weights $w_T(u)$ of every feature *u* in *T*:

$$W_T = \sum_{u \,\varepsilon\, T} w_T(u)\ .$$

To test the assumptions of the metaphor theories, *h* was defined as follows: The unweighted shared set *S* was compared with the weighted features in *Z*, rendering W_T . The features in *S* were not weighted, because otherwise it could never be the case that $W_Z = W_T$, which was the prediction of two models.

Similarly, the overlap was established between the unweighted S_{ll} , S_{lf} , S_{fl} , S_{ff} on the one hand and the weighted Z_l (the literal features in the interpretation) on the other. These comparisons rendered the values ${}_{Zl}W_{Tll}$, ${}_{Zl}W_{Tlf}$, ${}_{Zl}W_{Tfl}$, ${}_{Zl}W_{Tff}$. **For clearness' sake, ${}_{Zl}W_{Tll}$ means that W_T is the result of searching the weighted Z_l with the unweighted S_{ll}** . The unweighted S_{ll} , S_{lf} , S_{fl} , S_{ff} were also used to search the weighted Z_f (the figurative features in the interpretation), yielding the values ${}_{Zf}W_{Tll}$, ${}_{Zf}W_{Tlf}$, ${}_{Zf}W_{Tfl}$, ${}_{Zf}W_{Tff}$.

The unweighted S_{rr} , S_{xr} , S_{gr} were used to calculate the overlap with the weighted Z_r (the relations in the interpretation) and the weighted Z_g (grounds in the interpretation). The respective values are ${}_{Zr}W_{Trr}$, ${}_{Zr}W_{Txr}$, ${}_{Zr}W_{Tgr}$ and ${}_{Zg}W_{Trr}$, ${}_{Zg}W_{Txr}$, ${}_{Zg}W_{Tgr}$.

It should be recalled that the *expression* and *context* condition were preceded by a *single term* condition, in which no literal, figurative and relation distinctions were made. Recall also that one subject group worked in the sequence '*single term*, *expression*, *expression interpretation*' and the other in the sequence '*single term*, *context*, *interpretation in context*'.[9] Hence, those shared features S in *single term* that were also in Z are called T_1 . Those shared features S in *expression* or *context* that were also in Z are called T_2 .

For the test of the comparison model, two between-subject conditions (*expression interpretation* vs *interpretation in context*) were administered, each containing three expression types (within subjects), incorporating three weighted set sizes: (W_{T1} vs W_{T2} vs W_Z). *A*- and *B-term* differences were disregarded. MANOVAs were executed for the contrasts (W_{T1} vs W_{T2}), (W_{T1} vs W_Z) and (W_{T2} vs W_Z), relative to expression type and condition. Another MANOVA contrasted the expression types for W_Z in the two conditions (Table 5.6, Appendix). As mentioned above, literal, figurative and relation features were not distinguished in the preceding *single term* conditions, so that they were discarded for further analysis.

For the test of the anomaly and interaction model, MANOVAs (Table 5.6) were run for the two between-subject conditions (*expression interpretation* vs *interpretation in context*), containing three within-subject expression types with one of the following two set types:

(${}_{Zl}W_{T2ll}$ vs ${}_{Zf}W_{T2ll}$ or W_{Zl}), (${}_{Zf}W_{T2ll}$ vs W_{Zf})
(${}_{Zl}W_{T2lf}$ vs ${}_{Zf}W_{T2lf}$ or W_{Zl}), (${}_{Zf}W_{T2lf}$ vs W_{Zf})
(${}_{Zl}W_{T2fl}$ vs ${}_{Zf}W_{T2fl}$ or W_{Zl}), (${}_{Zf}W_{T2fl}$ vs W_{Zf})
(${}_{Zl}W_{T2ff}$ vs ${}_{Zf}W_{T2ff}$ or W_{Zl}), (${}_{Zf}W_{T2ff}$ vs W_{Zf}), (W_{Zl} vs W_{Zf})
(${}_{Zr}W_{T2rr}$ vs ${}_{Zg}W_{T2rr}$ or W_{Zr}), (${}_{Zg}W_{T2rr}$ vs W_{Zg})
(${}_{Zr}W_{T2xr}$ vs ${}_{Zg}W_{T2xr}$ or W_{Zr}), (${}_{Zg}W_{T2xr}$ vs W_{Zg})
(${}_{Zr}W_{T2gr}$ vs ${}_{Zg}W_{T2gr}$ or W_{Zr}), (${}_{Zg}W_{T2gr}$ vs W_{Zg}), (W_{Zr} vs W_{Zg}),

where ${}_{Zl}W_{T2ll}$ **means that** W_{T2} **is the result of searching the weighted** Z_l **with the unweighted** S_{ll} **in** ***expression interpretation*** **or** ***interpretation in context.***

Results of the shared sets retrieved in the interpretation sets Z

Table 5.5 shows the grand mean weighted feature set sizes of the interpretation sets Z and the set sizes of the weighted shared features retrieved in Z. Moreover, MANOVAs are displayed for the effects of condition, expression type and set type (Appendix Table 5.6).

Regarding the retrieved shared sets, Table 5.6 shows that:

(I) W_{T1} and W_{T2} were smaller than W_Z (main effect of set type: ❶$_{d}$). (II) More shared features were retrieved in Z for metaphors than for anomalies (main effect of expression type: ❷$_{d}$). (III) W_{T1} was smaller than W_{T2} , the difference being stronger for metaphors than for anomalies (interaction of set type by expression type: ❸$_{d}$).

With reference to literal and figurative features, it was found that:

(I) In *interpretation in context*, more figurative features (W_{Zf}) were activated than in *expression interpretation*. However, in the cases ${}_{Zf}W_{T2lf}$, ${}_{Zf}W_{T2fl}$ and ${}_{Zf}W_{T2ff}$, fewer shared features were retrieved in *interpretation in context*. Only for ${}_{Zf}W_{T2ll}$, there was a small increase (interaction of condition by set type: ❹$_{d}$). (II) More literal than figurative features were activated: $W_{Zl} > W_{Zf}$ (main effect of set type: ❺$_{d}$). (III) Purely literal overlap was better retrieved in the literal than in the figurative features of Z: ${}_{Zl}W_{T2ll} > {}_{Zf}W_{T2ll}$ (main effect of set type: ❻$_{d}$). (IV) The retrieved shared sets - either literal, figurative, or mixed - were significantly smaller than the applicable weighted set sizes W_Z (main effect of set type: ❼$_{d}$). (V) More features of the mixed shared set S_{lf} were retrieved in the Z-sets of metaphors than in those of literal expressions. In addition, more features of the mixed shared set S_{fl} were retrieved in the Z-sets of metaphors than in those of anomalies (main effect of expression type: ❽$_{d}$). (VI) Whereas the mixed shared set S_{lf} retrieved in Z_l (${}_{Zl}W_{T2lf}$) was smaller for literal expressions than for anomalies, W_{Zl} was larger for literal expressions. More shared features of S_{fl} were retrieved in *Zl* for literal expressions than for anomalies (${}_{Zl}W_{T2fl}$), while W_{Zl} was larger for literal expressions than for anomalies (interaction of set type by expression type: ❾$_{d}$).

Table 5.5: Grand means of the weighted feature set size W_Z (= W_{Zl} + W_{Zf}), W_{Zr} and W_{Zg} , and W_{T1} , W_{T2} , ${}_{Zl}W_{T2ll}$, ${}_{Zf}W_{T2ll}$, ${}_{Zl}W_{T2lf}$, ${}_{Zf}W_{T2lf}$, ${}_{Zl}W_{T2fl}$, ${}_{Zf}W_{T2fl}$, ${}_{Zl}W_{T2ff}$, ${}_{Zf}W_{T2ff}$, ${}_{Zr}W_{T2rr}$, ${}_{Zg}W_{T2rr}$, ${}_{Zr}W_{T2xr}$, ${}_{Zg}W_{T2xr}$, ${}_{Zr}W_{T2gr}$, ${}_{Zg}W_{T2gr}$.

Grand mean number of weighted features W_Z (= W_{Zl} + W_{Zf}), W_{Zr} and W_{Zg}

Condition	Literals		Metaphors		Anomalies	
	W_{Zl}	W_{Zf}	W_{Zl}	W_{Zf}	W_{Zl}	W_{Zf}
expression interpretation	4.90	1.28	4.23	2.53	2.70	2.31
interpretation in context	4.76	3.65	4.60	3.08	3.08	3.02
	W_{Zr}	W_{Zg}	W_{Zr}	W_{Zg}	W_{Zr}	W_{Zg}
expression interpretation	.804	.526	1.76	1.47	2.22	2.25
interpretation in context	2.58	2.68	3.20	3.32	2.72	2.15

Grand mean number of weighted shared features (W_T) retrieved in *Z*

Condition	Literals		Metaphors		Anomalies	
	W_{T1}	W_{T2}	W_{T1}	W_{T2}	W_{T1}	W_{T2}
expression interpretation	.030	.237	.045	.424	.022	.126
interpretation in context	.134	.190	.038	.355	.010	.090
	${}_{Zl}W_{T2ll}$	${}_{Zf}W_{T2ll}$	${}_{Zl}W_{T2ll}$	${}_{Zf}W_{T2ll}$	${}_{Zl}W_{T2ll}$	${}_{Zf}W_{T2ll}$
expression interpretation	.156	.000	.179	.017	.015	.004
interpretation in context	.110	.017	.124	.017	.048	.000
	${}_{Zl}W_{T2lf}$	${}_{Zf}W_{T2lf}$	${}_{Zl}W_{T2lf}$	${}_{Zf}W_{T2lf}$	${}_{Zl}W_{T2lf}$	${}_{Zf}W_{T2lf}$
expression interpretation	.015	.000	.045	.028	.022	.004
interpretation in context	.017	.003	.059	.021	.017	.007
	${}_{Zl}W_{T2fl}$	${}_{Zf}W_{T2fl}$	${}_{Zl}W_{T2fl}$	${}_{Zf}W_{T2fl}$	${}_{Zl}W_{T2fl}$	${}_{Zf}W_{T2fl}$
expression interpretation	.015	.000	.031	.052	.004	.011
interpretation in context	.017	.014	.069	.017	.010	.010
	${}_{Zl}W_{T2ff}$	${}_{Zf}W_{T2ff}$	${}_{Zl}W_{T2ff}$	${}_{Zf}W_{T2ff}$	${}_{Zl}W_{T2ff}$	${}_{Zf}W_{T2ff}$
expression interpretation	.007	.022	.041	.031	.000	.070
interpretation in context	.017	.021	.038	.048	.003	.007
	${}_{Zr}W_{T2rr}$	${}_{Zg}W_{T2rr}$	${}_{Zr}W_{T2rr}$	${}_{Zg}W_{T2rr}$	${}_{Zr}W_{T2rr}$	${}_{Zg}W_{T2rr}$
expression interpretation	.000	.011	.034	.003	.093	.048
interpretation in context	.017	.003	.034	.007	.038	.014
	${}_{Zr}W_{T2xr}$	${}_{Zg}W_{T2xr}$	${}_{Zr}W_{T2xr}$	${}_{Zg}W_{T2xr}$	${}_{Zr}W_{T2xr}$	${}_{Zg}W_{T2xr}$
expression interpretation	.000	.011	.034	.014	.022	.000
interpretation in context	.028	.007	.052	.041	.031	.010
	${}_{Zr}W_{T2gr}$	${}_{Zg}W_{T2gr}$	${}_{Zr}W_{T2gr}$	${}_{Zg}W_{T2gr}$	${}_{Zr}W_{T2gr}$	${}_{Zg}W_{T2gr}$
expression interpretation	.004	.000	.010	.000	.030	.052
interpretation in context	.007	.003	.010	.010	.017	.028

Turning to relations and grounds, the following results were found:

(I) In *interpretation in context*, more grounds and relations (W_{Zr}+W_{Zg}) were created than in *expression interpretation* (main effect of condition: ❿$_d$). (II) Although fewer relations (W_{Zr}) and grounds (W_{Zg}) were created in *expression interpretation* than in *interpretation in context*, more shared relations and grounds (${}_{Zr}W_{T2rr}$, ${}_{Zg}W_{T2rr}$, ${}_{Zg}W_{T2xr}$, ${}_{Zg}W_{T2gr}$) were retrieved in the former than in the latter condition (interaction of condition by set type: ❶$_e$). (III) In all cases, retrieved shared sets were smaller than the weighted *Z*-sets: ${}_{Zr}W_{T2rr}$ <

W_{Zr} , $_{Zg}W_{T2rr} < W_{Zg}$, $_{Zr}W_{T2xr} < W_{Zr}$, $_{Zg}W_{T2xr} < W_{Zg}$, $_{Zr}W_{T2gr} < W_{Zr}$, $_{Zg}W_{T2gr} < W_{Zg}$ (main effect of set type: ❷$_e$).

Discussion of the shared sets retrieved in Z

Above, the shared set S was found as ($X \cap Y$). This shared set was used as a file to search the interpretation sets Z for overlap ($S \cap Z$). The result of this search action was the retrieved shared set T. It was argued that if all features in S were found again in Z, the size of the weighted set T should be equal to the weighted set size of Z. If Z is (much) larger than T, interpretation should not be considered identical to finding the shared set.

The comparison model states that the shared set is identical to the interpretation set Z, irrespective of context. In other words, no condition effects should be found. If the shared set S is exactly the same as Z, comparisons between the values (W_{T1} , W_{T2}) and W_Z should show no differences. Thus, effects involving set type should be absent, so that only the main effect of expression type remains.

Indeed, no effects involving condition occurred on W_Z , nor on W_{T1} or W_{T2} (Appendix Table 5.6). For W_Z , no other effects were significant either. The sizes of the Z-sets did not differ significantly, so that expression type differences did not show up in the activation of interpretation features. This is not expected, because the shared set - and thus the Z-sets - of literal expressions should be larger than for metaphors, which should be larger than for anomalies. Since the comparison model assumes that Z is identical to the shared set, this ordinal pattern should also have been found for Z-sets.

Nevertheless, for the retrieved shared sets ($W_{T1} + W_{T2}$), the main effect of expression type did show a pattern in line with the comparison model. In the Z-sets of metaphors, more shared features were reactivated than in those of anomalies. The difference with literal Z-sets, however, was lacking.

The main effect of set type shows that W_{T1} was smaller than W_{T2} . According to the interaction of set type by expression type, this difference was larger for metaphors than for anomalies. Moreover, the main effect of set type shows that W_{T1} and W_{T2} were smaller than W_Z . These results are not expected if Z consisted only of (reactivated) shared features. In that case, the main effect of set type would show that $W_{T1} = W_{T2} = W_Z$, with p-values close to 1. Since the opposite occured ($p = .000$), **interpretation was not at all identical to finding the shared set. The retrieved shared set size of metaphors did not differ from that of literal expressions, but only from that of anomalies**.

The anomaly model posits that the shared literal features are the Z-sets of literal expressions, and the shared figurative features are the Z-sets of metaphors. Again, context has no effect, so that condition effects should be absent. If the set of shared literal features S_{ll} is identical to Zl in literal expressions, and if S_{ff} is identical to Zf in metaphors, then ${}_{Zl}W_{T2ll} = W_{Zl}$ in literal expressions, and ${}_{Zf}W_{T2ff} = W_{Zf}$ in metaphors. Effects of set type should be absent, and the main effect of expression type should be present.

However, condition effects did occur. The main effect of condition and the interaction of condition by set type indicate that in *interpretation in context*, more figurative features were activated than in *expression interpretation*.

The main effect of set type shows that more literal than figurative features were activated. Additionally, purely literal overlap was better retrieved in the literal than in the figurative features of *Z*. More importantly, the retrieved shared sets - either literal, figurative, or mixed - were significantly smaller than the set sizes of the applicable *Z*-sets. These results are not expected if the *Z*-sets of literal expressions exclusively consisted of shared literal features, and the *Z*-sets of metaphors of shared figurative features. Again, the p-values were close to zero, so that the results also militate against the anomaly view on interpretation.

The main effect of expression type supports the presumption that larger shared sets are employed to interpret metaphors than anomalies. More features of the mixed shared set S_{fl} were retrieved in *Z*-sets of metaphors than of anomalies. In comparison with literal expressions, metaphors also re-activated more features of the mixed shared set S_{lf} in *Z*. This is congruous with the anomaly model, if the notion of purely figurative overlap is abandoned. Positive differences for literal expressions compared with anomalies were found in shared sets retrieved in literal features of *Z*. Only for S_{lf} was this difference reversed in the interactions of set type by expression type: $W_{Zl}(\mathrm{L} > \mathrm{A})$ vs ${}_{Zl}W_{T2lf}(\mathrm{L} < \mathrm{A})$, which is not a prediction of the anomaly model.

Again, **interpretation was not at all identical to finding the shared set, either literal, figurative or mixed. Metaphors were different from anomalies in that more features of the mixed shared set S_{fl} were retrieved in the interpretation. Metaphors were different from literal expressions in that more features of the mixed shared set S_{lf} were retrieved in the interpretation.**

The interaction model implies that the *Z*-sets of metaphors are based on relations, whereas the *Z*-sets of other expression types are not. Context is important for the interpretation, so that effects involving expression type should show more relation overlap for *Z*-sets of metaphors than of other expression types. Adding context should increase this effect. If shared relations and grounds are fully responsible for the interpretation, effects of set type are missing.

In fact, the main effects of condition indicate that in *interpretation in context*, more grounds and relations were activated than in *expression interpretation*. However, the interactions of condition by set type show that more shared relations and grounds were retrieved in *expression interpretation* than in *interpretation in context*. In *expression interpretation*, shared relations and grounds contributed more to the interpretation than in *interpretation in context*. This cannot be the case, if the retrieved shared sets are identical to the *Z*-sets. The main effects of set type also illustrate the inequality between retrieved shared sets and *Z*-sets, which were always larger.

Effects involving expression type did not occur, which is in line with the finding of Table 5.3 (Appendix), that relation creation is sensitive to condition, but not to expression type. Since the interaction model advances that in metaphor interpretation, more relations and grounds are connected than in literals and anomalies, the absence of effects for expression type refutes the theory. Thus, **interpretation was not at all identical to connecting relations and/or grounds, and did not distinguish metaphors from other expression types**.

Expression type differences in interpretation were found in the retrieved shared sets, not in the interpretation set sizes. The features of the shared set *S* retrieved in *Z* could distinguish positively metaphors from anomalies. However, S_{fl} retrieved in *Z* could do just that, while S_{lf} could distinguish metaphors from literals. Moreover, literal expressions were positively distinguished from anomalies by the shared sets retrieved in literal features of *Z*: $({}_{Zl}W_{T2ll}+W_{Zl}\ ,\ {}_{Zl}W_{T2fl}+W_{Zl}\ ,\ {}_{Zl}W_{T2ff}+W_{Zl})$. As shown earlier in Table 5.4 (Appendix), the mixed shared feature sets S_{lf} and S_{fl} were the best predictors of expression type differences, and again, proved to be the best predictors of interpretation differences as well. The shared sets defined by the interaction model did not differentiate expression types, which was also inferred by Table 5.4.

In conclusion, **finding shared features was not equal to interpretation, irrespective of feature type. The anomaly model defined the best shared sets to distinguish the expression types and their interpretations**.

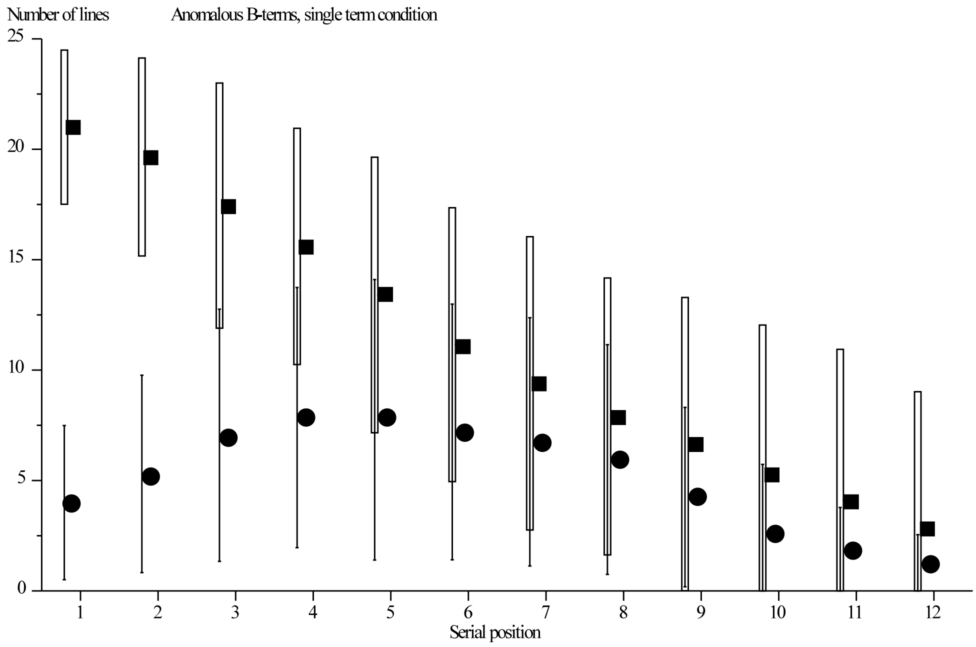
Serial positions of feature activation

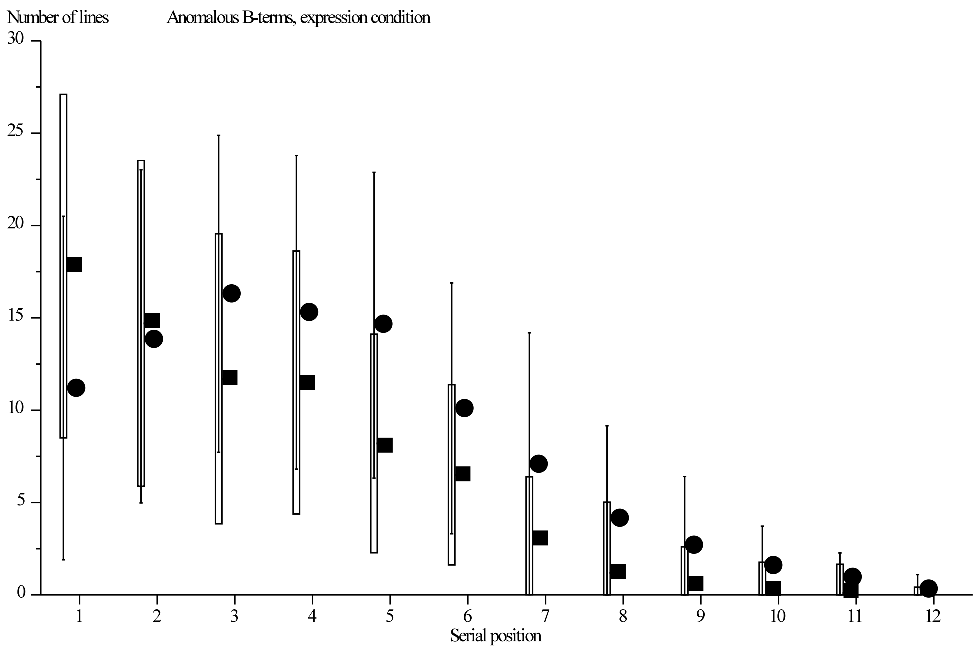
Literal and figurative features at serial positions

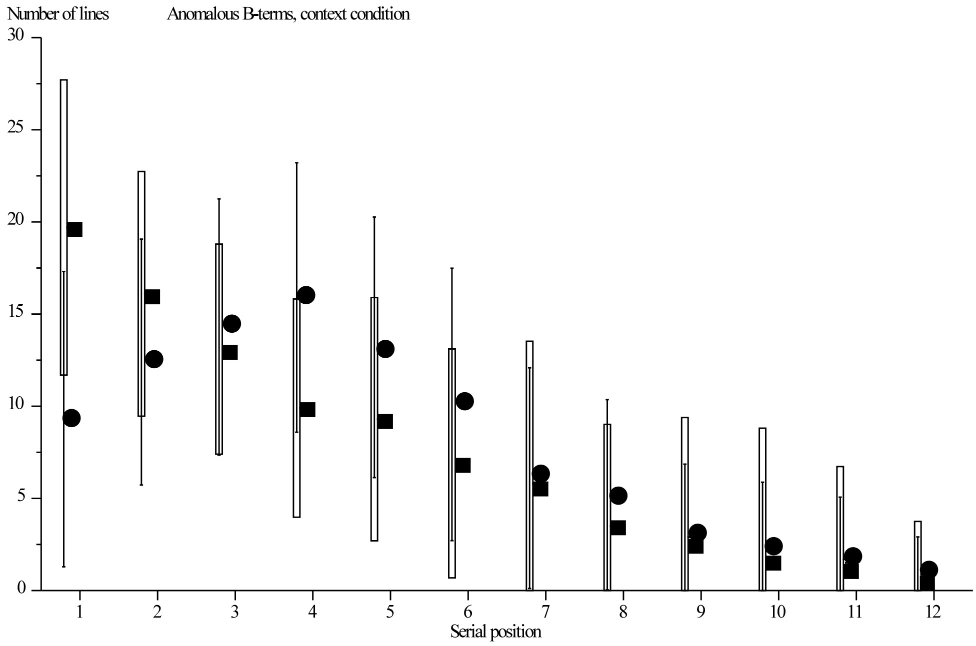
Anomaly theory claims that figurative features are activated when subjects run out of literal features. In that case, the feature lists should have more literal features in the beginning of a list, followed by a clear shift to figurative features. In particular, metaphors and anomalies should show this shift, whereas literal expressions may activate literal features only.

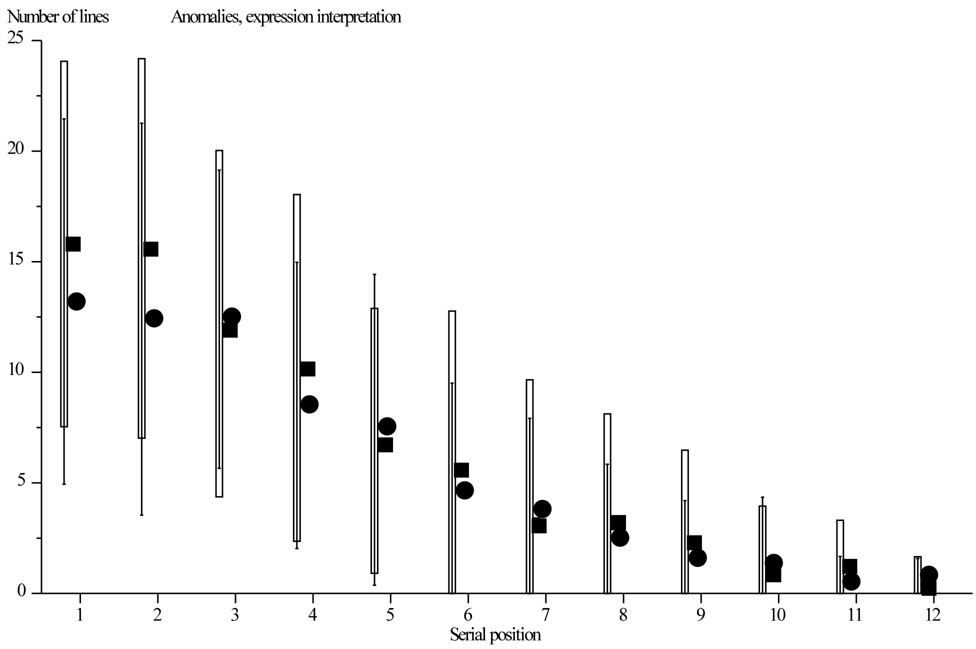
To test this deduction, the literal and figurative features were counted per line, and averaged over subjects for each literal, metaphoric and anomalous *A*- and *B-term* in each condition. The same happened for the interpretations. Except for anomalies in *expression* and *context*, the results of this analysis showed a pattern as illustrated by the anomalous *B-terms* in *single term* (Figure 5.15, upper panel).

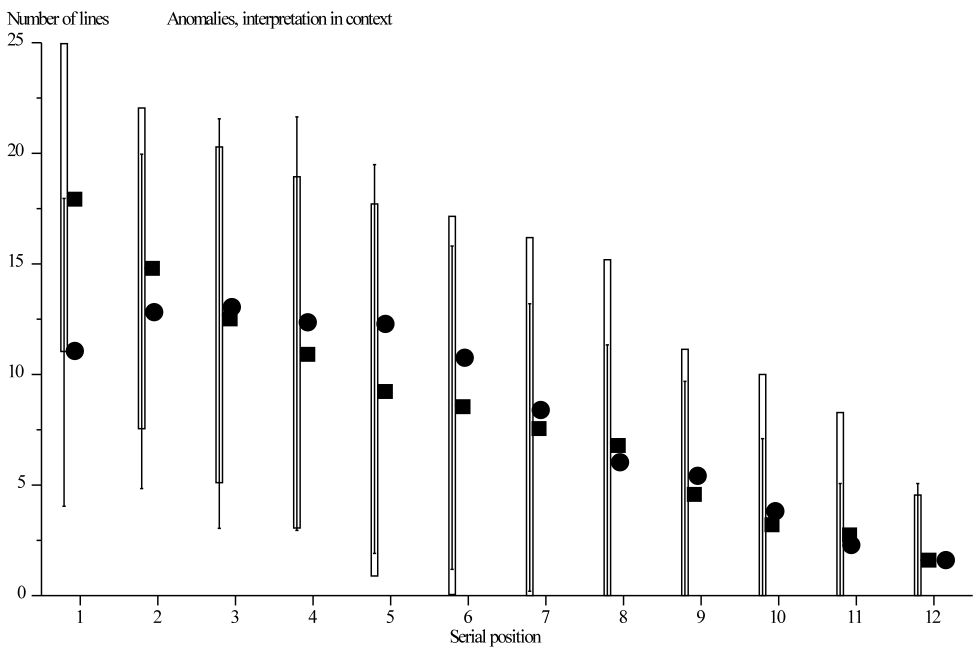
Figure 5.15: Grand mean literal (■) and figurative (●) features per line in the feature lists of the anomalous *B-terms* and anomaly interpretations. All other term types and interpretations in all conditions showed the pattern of the anomalous *B-terms* in *single term*. Standard deviations for literal features are demarcated by rectangulars, and for figurative features by solid lines.




Number of lines
Anomalous B-terms, expression condition
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
Serial position


Number of lines
Anomalous B-terms, context condition
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
Serial position


Number of lines
Anomalies, expression interpretation
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12
Serial position


Number of lines
Anomalies, interpretation in context
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12
Serial position

Note that the decline of the grand means in Figure 5.15 is due to the fact that not every subject typed as many lines, so that the mean number of lines in first position is larger than the mean number of lines in the twelfth position. Figure 5.15 shows that for anomalous *B-terms* in *single term* (and for all other term types in all other conditions), literal features were more frequent than figurative features at all positions. This pattern was found also for the *A-terms* (independent of expression type), literal and metaphoric *B-terms* and interpretations of literals and metaphors.

These results repudiate the anomaly model for two reasons. First, literal and figurative features appear to be activated together, not after each other. Although figurative features were less frequent, they were evenly distributed among positions, even for literal expressions. Second, the shift to higher frequencies of figurative features failed for metaphors in all conditions, and for anomalies in *single term*.

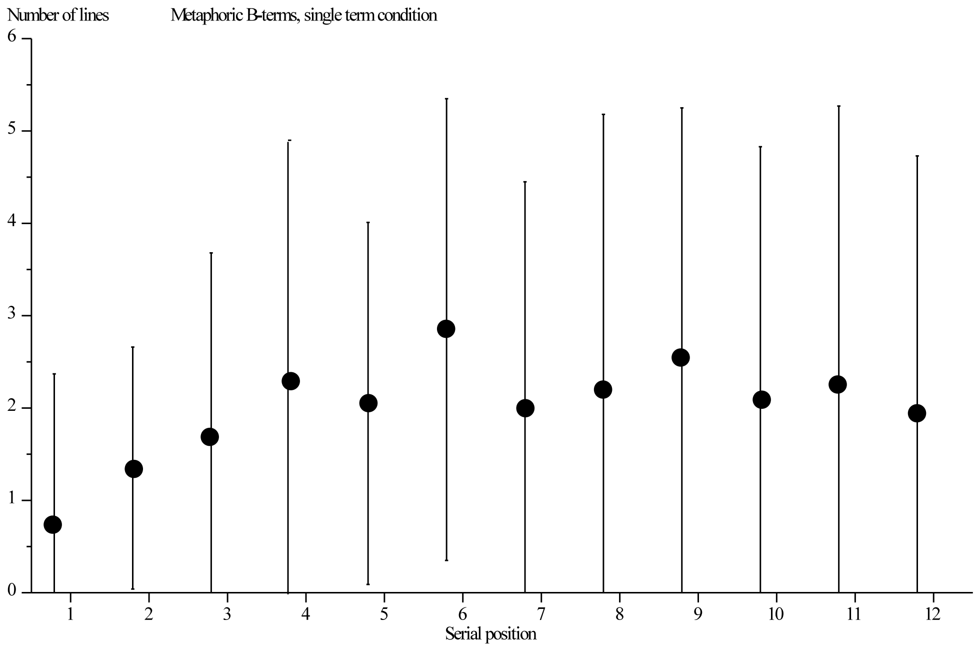
The rise of figurative features between position 3 and 6 seen for every term type exceeded the literal features only for anomalous *B-terms* in *expression* and *context*, and for the anomaly interpretations. Here, the figurative shift did occur, thus corroborating the anomaly model. In each case, **literal features were dominant in the initial activation, whereas the figurative shift was dependent of expression type (anomalous *B-terms*) and condition *(expression* and *context*).**

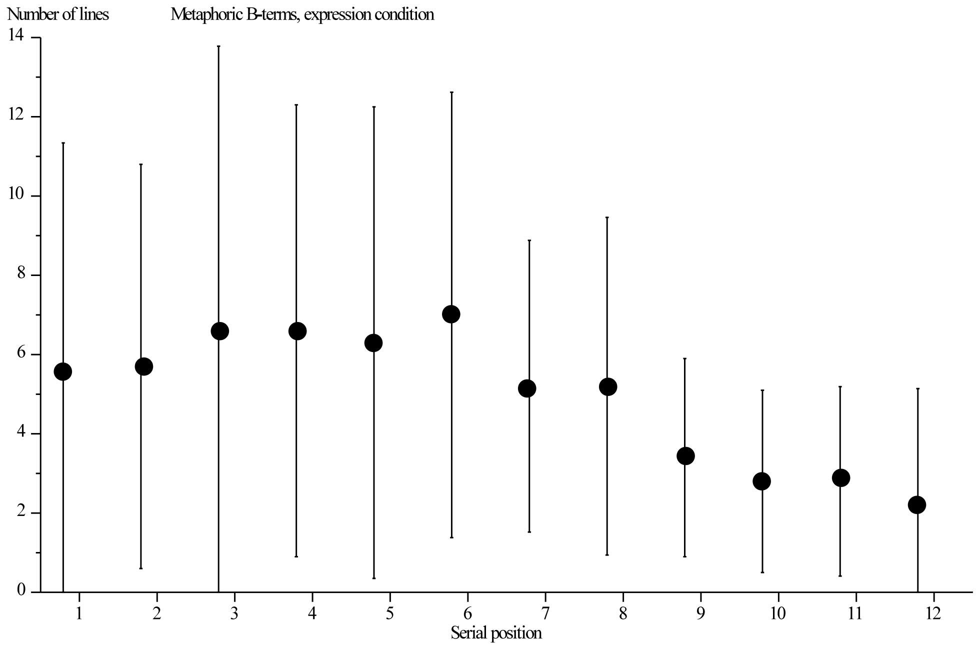
Relations at serial positions

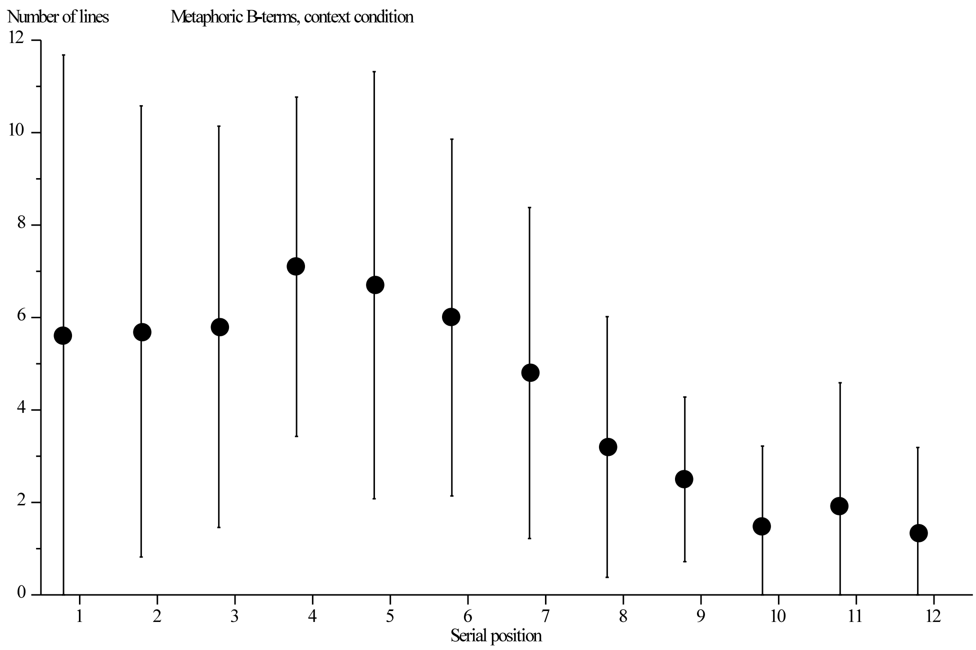
The interaction model predicts that relations are activated after the normal features are depleted. This is mainly a trait of the metaphoric and anomalous *B-terms*, dependent on the surrounding text. Therefore, relations should not be found in *single term*, not for *A-terms*, not for literal expressions, and not at the first positions in the feature lists. For metaphoric and anomalous *B-terms* in *expression* and *context*, a clear augmentation of relations should occur at second or following positions.

For each *A-* and *B-term*, then, in each expression type and condition, and for each interpretation, the relations were counted per line, and averaged over subjects. The grand means for metaphoric *B-terms* and metaphor interpretations are shown in Figure 5.16. All other term types and interpretations in all conditions showed similar results.

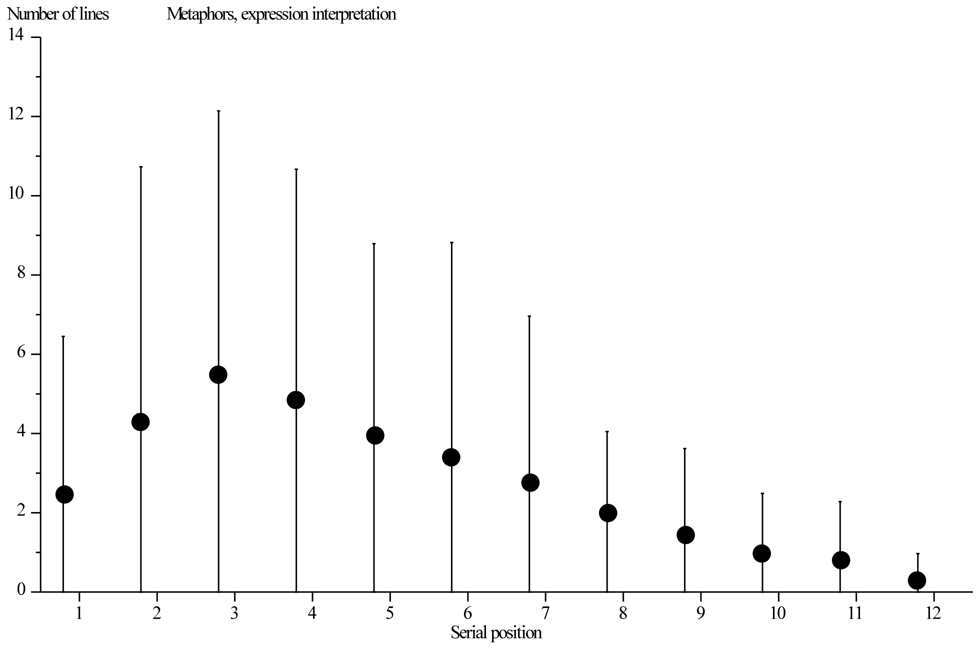
Figure 5.16: Grand mean number of relations (●) and standard deviations per line in the feature lists of the metaphoric *B-terms* and metaphor interpretations. All other term types and interpretations in all conditions showed comparable patterns.



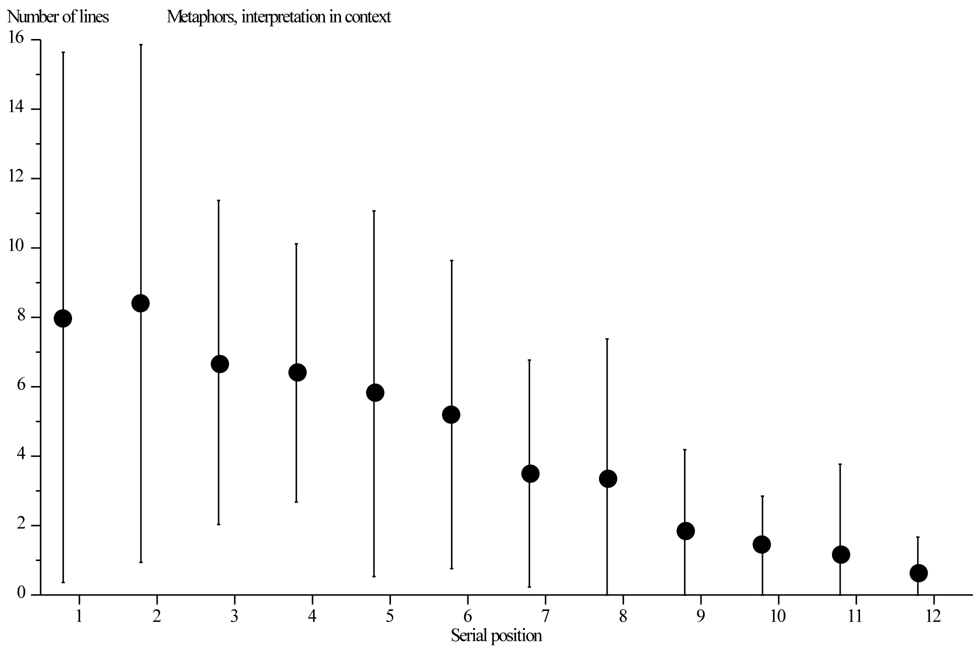




Number of lines
Metaphoric B-terms, context condition
0
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11
12
Serial position



Number of lines
Metaphors, expression interpretation
0
2
4
6
8
10
12
14
1
2
3
4
5
6
7
8
9
10
11
12
Serial position



Indeed, Figure 5.16 indicates that metaphoric *B-terms* rendered more relations at the third to sixth positions than at the preceding ones. The same results applied to anomalous *B-terms*, but also to all other term types (including *A-terms*), to literal expressions, and in all conditions, including *single term*. In other words, the interaction model is right about **the delayed activation of relations. However, this was not unique for metaphors, and developed independent of context.**

Check 1: Asymmetry differences among expression types

Asymmetry of feature sets?

The next issue is the transition from Tversky's focusing hypothesis to feature sets. Tversky claimed that the *A-term* is the focus of a comparison, because it is less salient. As an alternative option, it was suggested that the *A-term* was feature-poor. It was also suggested that different frequencies of features in the shared set may explain the preference in focus. The frequencies of shared features of the *A-term* set (${}_XW_S$) may be lower than the frequencies in the *B-term* set (${}_YW_S$). Asymmetry in similarity judgements, then, could be explained from differences *within* the shared set.

Knowledge about differences in focusing among expression types is important, because asymmetry may be part of the differences in processing. Thus, differences in focusing may influence reaction times, if a focused *B-term* is processed 'deeper' - and presumably slower - than a nonfocused *B-term*.

In Table 4.1 (Chapter 4), it was found that subjects almost unanimously chose the *A-term* as the focus of the expressions. Thus, *A-terms* should have smaller (weighted) feature sets than *B-terms* ($W_X < W_Y$), or the frequencies of the shared features should be less for the *A-term* set (${}_XW_S < {}_YW_S$).

However, Tables 5.3 and 5.4 indicate that the only cases in which a difference in the expected direction occurred were the second-order interactions on (W_X , W_Y) and (W_{Xa} , W_{Ya}), the interactions of term by expression type on (W_{Xl} , W_{Yl}). Metaphors in *single term* and *expression* had feature rich *B-terms*: M($W_X < W_Y$), M($W_{Xa} < W_{Ya}$). Metaphors were also sensitive to the interaction of term by expression type on literal feature sets: M($W_{Xl} < W_{Yl}$).

For the shared set, differences into the expected direction were found in the interaction of condition by term on W_S , WS_{xr} , WS_{gr} . In all conditions, it was found that ${}_XW_{Sx} < {}_YW_{Sr}$, and in *context* and *expression*, that ${}_XW_{Sg} < {}_YW_{Sr}$. *Expression* also showed that ${}_XW_S < {}_YW_S$. In all other cases, *A-* and *B-term* yielded equal or inverted ratios.

In conclusion, **the number of features for the *B-term* were only occasionally higher than for the *A-term*. Metaphors were the only expression type that occasionally showed more features for the *B-term*. Otherwise, feature sets for the *A-term* were often larger than for the *B-term*. There were virtually no differences in frequency of features in the shared set, so that the shared sets were usually symmetrical.**

Asymmetry in judgements of similarity and figurativeness?

These results are at odds with the findings of Tversky (1977) and Malgady & Johnson (1980). It could be that the stimulus materials in the present study differ from those of Tversky and Malgady & Johnson. Alternatively, feature set size or frequency of features in the shared set might not be the proper measures to establish asymmetry. To approach the problem from a different angle, asymmetry was directly estimated from similarity judgements.

An experiment was administered on estimates of similarity and figurativeness. Malgady & Johnson (1980) found that in similarity judgements, subjects focused on the nonsalient *A-term* and in judgements of figurativeness on the salient *B-term*. If literals are primarily judged for similarity and metaphors for similarity *and* figurativeness (from the first stage to the second), the focus should switch in metaphors, whereas in literal expressions it should not.

Four groups of 13 subjects served as paid volunteers in a rating task. They were undergraduates of language and literature, Dutch native speakers, aged between 18 and 27 years old. Subjects estimated the similarity and figurativeness of the expression types, either from 'the first term to the second' (*A* to *B*) or 'from the second term to the first' (*B* to *A*), either in *expression* or in *context*. Thus, four between-subject cells (*expression* '*A* to *B*', *expression* '*B* to *A*', *context* '*A* to *B*', *context* '*B* to *A*') comprised three within-subject expression types, each incorporating two dimensions (similarity vs figurativeness).

Similarity was defined as the extent to which the two stimuli were alike in outer appearance and literal meaning. Figurativeness was defined as the symbolic or nonliteral correspondence between the stimuli. Subjects used a 7-points scale, where 1 was specified as 'no resemblance', 7 as 'complete resemblance', and 4 as 'neutral'. They were unaware of the origin or status of the expressions, nor did they know the aim of the study. Prior to the experiment, subjects worked on nine training stimuli unconnected to the test set. Trials were mixed in a pseudo-random order. Variants of an expression were not presented in succession, and instances of the same expression type did not occur more than twice in a row. Literal, metaphoric and anomalous variants appeared first, second and last in their relative order of presentation, counterbalanced over expression types. Each subject received a different order of presentation. The grand mean scores on similarity and figurativeness and the results of MANOVA are displayed by Table 5.7 and Table 5.8 (Appendix).

Table 5.7: Grand mean scores of similarity (*SIM*) and figurativeness (*FIG*) for judgements from '*A* to *B*' and '*B* to *A*' of literals (L), metaphors (M) and anomalies (A) in *expression* and *context* (N = 52).

Grand mean similarity and figurativeness

		dimension					
		similarity			figurativeness		
Condition	direction	L	M	A	L	M	A
expression	*A* to *B*	5.22	2.12	1.48	4.03	4.53	2.69
context	*A* to *B*	5.02	2.30	1.37	4.15	4.82	2.84
expression	*B* to *A*	4.05	2.09	1.51	3.71	4.58	3.01
context	*B* to *A*	3.98	2.36	1.49	3.90	3.89	2.28

The interaction of dimension by expression type in Table 5.8 (Appendix) shows that:

(I) Scores were higher for similarity than for figurativeness, mainly due to the literal expressions. Similarity was higher for literals than for metaphors, which was higher than for anomalies. On the other hand, metaphors had the highest figurativeness, anomalies the lowest, and literal expressions were in between. (II) No effects of judgement direction (*A* to *B*, *B* to *A*) were significant.

Since the effects of judgement direction were statistically unreliable, asymmetry was not established. This finding is in line with the absense of differences between the frequencies of features in the shared set, whereas it is at odds with the finding that *A-terms* often had larger feature sets than *B-terms*. Thus, **Tversky's focusing hypothesis and the switch of focus proposed by Malgady & Johnson were not confirmed with the present stimulus materials**.

Visual inspection of the grand means, however, suggests that in *expression*, similarity in metaphors was highest from *A* to *B* and figurativeness from *B* to *A*. Although these differences were minor, metaphors were the only expression type that showed a trend in line with Malgady & Johnson. This trend occurred only in *expression*, which was the same condition as Malgady & Johnson employed. In other words, the switch of focus to the *B-term* for judgements of figurativeness may exclusively apply to metaphors, and may depend on condition. The higher similarity from *A* to *B* for metaphors in *expression* agreed with the finding that the feature sets for metaphoric *B-terms* were larger than for *A-terms* in this condition. Nevertheless, the asymmetry derived from the direct subject estimates was statistically unreliable.

In brief, **asymmetry played no fundamental role in processing the present stimulus set.** The absence of asymmetry may be due to the fact that the present study did not utilize asymmetry cues, and did not reverse noun order. However, judgements were contrasted from the first term to the second with those from the second to the first. In the earlier work, stimuli were constructed with inherent asymmetric cues, such as 'the son looks *more* like the father' versus 'the father looks *more* like the son' (Tversky 1977). Another manipulation reversed the nouns in subject and object position, as in the above example and in 'mountains are kings' versus 'kings are mountains' (Malgady & Johnson 1980) or 'lectures are like sleeping pills' versus 'sleeping pills are like lectures' (Ortony, Vondruska, Foss & Jones 1985).

Cues for asymmetry may actually impose asymmetric judgements that are not obtained without those cues. Again, judgements from the first term to the second may induce asymmetry when uncontrolled by judgements from the second term to the first. Moreover, reversing the nouns changes sentence meaning, so that the stimuli in a pair may be incomparable with respect to their asymmetrical effects. 'Mountains are kings' does not mean the same as 'kings are mountains', so that asymmetry is confounded with change in meaning.

This is in line with the argument of Gleitman, Gleitman, Miller & Ostrin (1996). They reproduced the asymmetry effect for words such as 'similar to', but also for 'identical to' and 'equal to'. Nevertheless, they argued that '*A* is similar to *B*' has a different meaning than '*B* is similar to *A*'. Both are symmetric statements on similarity, namely 'with respect to'. 'North Korea is similar to Red China with respect to communism' is as symmetric as 'Red China is similar to North Korea with respect to political fragmentation'. Leaving this implicit context out, makes them two different sentences which seem comparable, but in fact are not. Experimental evidence showed that the perception of asymmetry - among others - was dependent on the position of the nouns in the sentence. For instance, asymmetry was found in comparisons with nonsense words: 'The ZUM is identical to the GAX' vs 'the GAX is identical to the ZUM' (cf. note 2).

In addition, Tversky and Malgady & Johnson investigated only one expression type: Similes and metaphors, respectively. It may be that mixing with other expression types (literal expressions and anomalies) affects the asymmetry. For instance, if the additional expression types are asymmetric in the opposite direction, the asymmetry of metaphors may be drowned by an increase of noise.

Moreover, the role of context in Malgady & Johnson was quite clear. By adding or deleting an adjective ('[red] blood poors like rain'), it was known fairly precisely which feature ('red') was emphasized for which term ('blood'). Malgady & Johnson reported that when the adjective was shared by both terms, similarity and figurativeness were increased. When the adjective emphasized a distinctive feature of the (nonsalient) *A-term*, similarity was increased, whereas figurativeness decreased. Adjectives that highlighted a distinctive feature of the (salient) *B-term* decreased similarity and increased figurativeness. In the present study, however, the effects of context were more diffuse, and may not have influenced the direction of judgement as much as the adjectives of Malgady & Johnson did.

In conclusion, **asymmetry may be an unstable effect, only induced by specific cues, and limited to similes and metaphors when not mixed with other expression types. When other factors are added, variability may be so severe, that effects fail to be significant. The present stimulus set did not suffer from systematic asymmetry effects**.

Check 2: The repetition effect

The feature elicitation experiment in Section 5.4 confounded effects of term repetition and condition. Terms were not repeated in *single term*, while they were in *expression* and *context*. A test was designed to control for the effect of repeating terms in *expression* and *context*.

It should be recalled that the subject groups, working in either *expression* (N = 10) or *context* (N = 10), started with a *single term* condition. Certain terms (e.g., 'death', 'heart') were repeated more frequently, whereas others (e.g., 'wall', 'stone') were repeated only once, i.e. the first presentation was in *single term* and the second in *expression* or *context*.

Hence, potential differences in feature rates from the second presentation on may be the result of repetitions within a condition. It is possible that the smaller feature set size in *expression* and *context* was due to specific practice on the terms. It could be that the number of features declines as a term is seen more often.

The number of weighted features (words) was counted for each repeated term per subject in *expression* and *context*, and contrasted with the number of weighted features produced for that term in the preceding *single term*. For each repetition, the grand means per expression type were calculated (Table 5.9). The analysis was limited to three repetitions, so as to reduce the number of pairwise comparisons.

Thus, the analyzed terms were presented four times (presentation 1, 2, 3, 4), and six pairwise comparisons were tested by MANOVA. The set-up combined two between-subject conditions ((*single term* and *expression*) versus (*single term* and *context*)) with three within-subject levels for expression type, and two sub-levels within each level, namely, one of the six presentation pairs: 1-2, 1-3, 1-4, 2-3, 2-4, 3-4. Effects without presentation pair as factor are not of interest to the analysis, so that the main effects of condition, expression type, as well as the interactions between condition and expression type, will not be listed in Table 5.10 (Appendix).

Table 5.9: The grand mean number of weighted features (words) per term presentation. Presentation 1 was performed in the preceding *single term*, whereas presentation 2 up to 4 were the repetitions in either *expression* or *context*.

Grand mean number of features per term

Condition 1	Presentation	Literals	Metaphors	Anomalies
single term	1	10.61	12.63	11.40
expression	2	9.01	10.37	7.06
expression	3	9.15	10.57	8.00
expression	4	9.20	10.65	7.80
Condition 2				
single term	1	12.91	13.22	11.93
context	2	12.35	10.67	8.83
context	3	12.45	10.51	8.53
context	4	12.40	10.00	8.20

Observe in Table 5.10 (Appendix), that:

(I) The main effect of presentation was significant for the difference between the first and all further presentations ($①_f$). For the differences between the second up to fourth presentation, no main effect was found. (II) There was no systematic interaction between condition and presentation, except for presentation 2 vs 4 at a liberal α-level. In condition 1, presentation 2 yielded less features than presentation 4, reversing the effect in condition 2 ($②_f$). (III) The significant interaction of presentation with expression type indicated that L(1 > 2), whereas A(1 >> 2) ($③_f$). At a liberal α-level (.05 > p > .0116), L(1 > 4), whereas M(1 >> 4), and A(1 >> 4) ($④_f$). No significant higher-order interactions were found for any presentation pair.

The main effect of presentation corroborates that there was a difference between the number of elicited features for the first presentation of a term in *single term* and its first repetition in *expression* or *context*. The number of weighted features for repetitions of these terms, however, did not differ within a condition. The direction of the main effect suggests that reduction of features took place, rather than accumulation. When context was added, the only exception to this pattern was that more features were found for presentation 4 than for 2 ($②_f$).

Thus, **the first presentation of a term yielded more features than the next presentations, which may be interpreted as a condition effect, because all effects of presentation within *expression* and *context* were insignificant**. The only exception is that at a liberal α-level, presentation 4 produced more features than presentation 2 in *context*. However, **the check is not water tight, because the repetition effect might be entirely situated between the first and second presentation**. A better control would have been to repeat each term several times in *single term*.

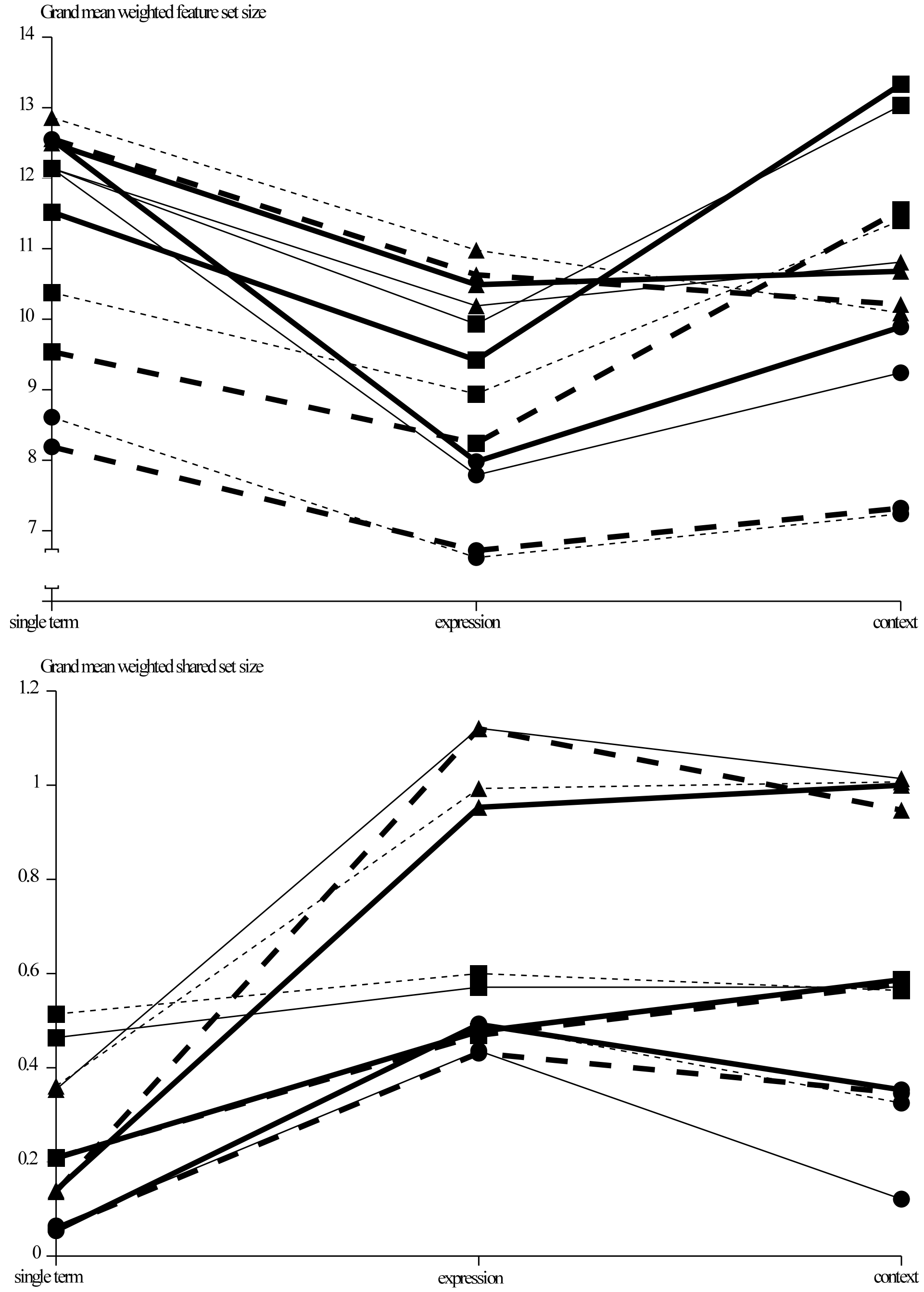
Check 3: Word frequency effect

Different word frequencies may entail different numbers of elicited features. An infrequent word indicates that it is rarely used, so that its reference may be unfamiliar. People know less about unfamiliar things, so that the number of associated features may be small for infrequent words.

In Chapter 4, Table 4.2 (Chapter 4) showed that anomalous *B-terms* were a low word frequency group. It may be, then, that the low feature rates for anomalous *B-terms* (e.g., Figure 5.0) are predicated by the low word frequencies. Thus, word frequency is a covariate of the stimuli. However, an F2 or quasi-F can only be calculated while using univariate tests. Unfortunately, SPSS only furnishes a classic multivariate model, in which varying covariances cannot be estimated. Therefore, it can only calculate covariances on the subject level (for instance, for age or IQ), not on the stimulus level.

To survey yet the effects of word frequency, an extra within-factor was constructed, namely word frequency (high-frequent vs low-frequent) within term. Terms within a term type (*A-term*, literal, metaphoric or anomalous *B-term*) were ranked according to their word frequencies, and then split half, resulting in two groups per term type: A high- and a low-frequent group relative to the term type. MANOVA was performed on this 3*3*2*2 design, and it was expected that the interactions of expression type by term by word frequency would show small feature set sizes and small contributions to the shared set for anomalous *B-terms* with low word frequencies compared to those with high word frequencies. Weighted feature set size (W_X , W_Y) and weighted shared set size (W_S) served as the dependent variables (Figure 5.17).

Figure 5.17: Grand mean weighted feature set size (upper panel) and weighted shared set size (lower panel) for condition, expression type, term and word frequency. ■ = literal expressions, ▲ = metaphors, ● = anomalies. Solid lines refer to set *X*, dashed to *Y*. Bold lines are for high-frequent terms, thin lines for low-frequent terms. Measure M2.



Effects in Table 5.11 (Appendix) are indexed by white circled figures with subscript (g), and they show the following:

(I) High-frequent *B-terms* activated smaller feature sets than low-frequent *B-terms*; for *A-terms*, the effect was the opposite (interaction of term by word frequency ①$_g$). (II) In literal expressions, high-frequent terms elicited fewer features than low-frequent ones; for metaphors, the ratios were equal. For anomalies, however, high-frequent terms elicited more features than low-frequent ones (interaction of expression type by word frequency ②$_g$). (III) In *single term*, high-frequent terms produced fewer shared features than low-frequent terms, whereas this difference reversed in *context* (interaction of condition by word frequency ③$_g$). (IV) In literal expressions and metaphors, high-frequent terms elicited fewer shared features than low-frequent ones. By contrast, high-frequent terms in anomalies elicited more shared features than low-frequent terms (interaction of expression type by word frequency ④$_g$).

First of all, it should be noted that none of the interactions of expression type by term by word frequency were significant on weighted feature set size or weighted shared set size. Thus, anomalous *B-terms* with low word frequencies did not negatively affect (shared) feature rates, although this was expected. Second, most of the word frequency effects on feature elicitation were chaotic: The sign between ratios reversed, often indicating higher scores for low-frequent than for high-frequent terms.

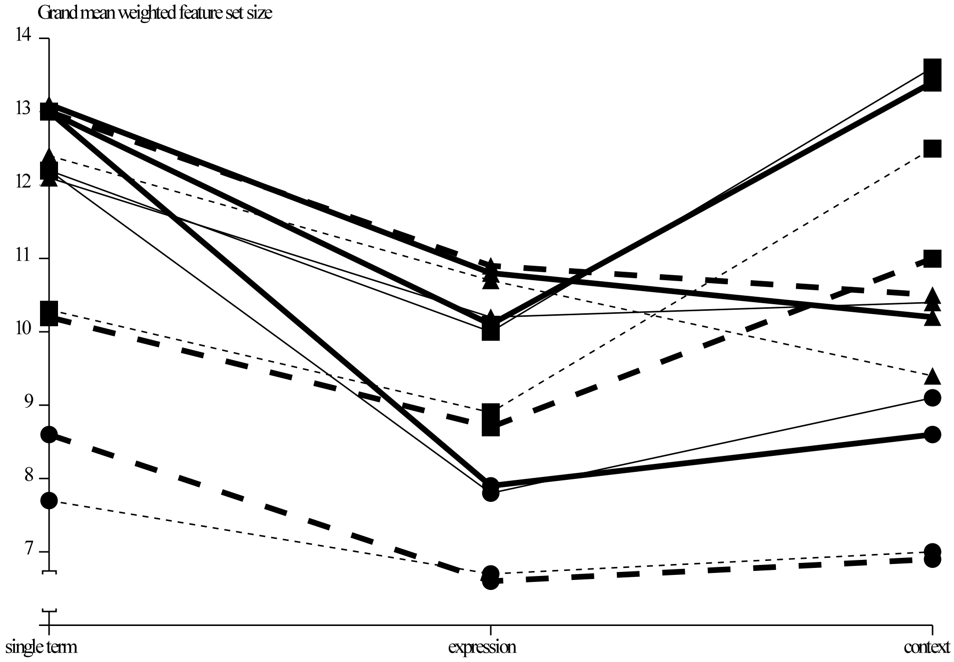
The only effects that may indicate lower scores for low-frequent terms in anomalies are found in the interactions of expression type by word frequency (②$_g$ and ④$_g$). However, these are not effects due to the low-frequent anomalous *B-terms*. The grand mean weighted feature set size and weighted shared set size show that - across conditions - low-frequent anomalous *B-terms* evoke more (shared) features than high-frequent ones. The lower scores for anomalies, thus are the effect of the high-frequent anomalous *B-terms* and the low-frequent *A-terms* in anomalies.

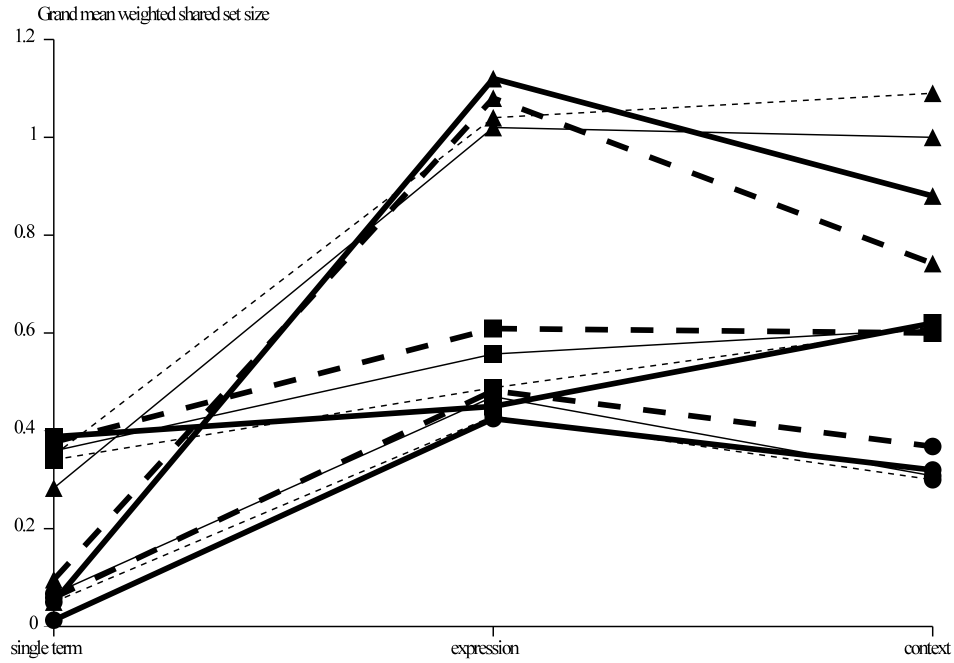
In sum, **word frequency may affect feature elicitation, but only in a random way. The expected negative effect for low word frequency of anomalous *B-terms* was not substantiated. The lower-frequent anomalous *B-terms* rendered larger (shared) set sizes than the higher-frequent ones. Word frequency should be considered noise.**

Check 4: Lexical ambiguity effect

Returning to the issue of lexical ambiguity, Table 4.3 (Chapter 4) listed all ambiguous terms in the expression set of Table 4.1. Lexical ambiguity, it will be recalled, may influence the number of elicited features, because features from multiple meanings might be produced. Lexical ambiguity also is a covariate of the stimuli. The availability of more meanings might increase feature set size, as well as the size of the shared set. To examine potential effects of lexical ambiguity, an

Figure 5.18: Grand mean weighted feature set size (upper panel) and weighted shared set size (lower panel) for condition, expression type, term and lexical ambiguity. ■ = literal expressions, ▲ = metaphors, ● = anomalies. Solid lines refer to set *X*, dashed to *Y*. Bold lines are for ambiguous terms, thin lines for nonambiguous terms. Measure M2.





extra within-factor was added, namely ambiguity (ambiguous vs nonambiguous) within term. MANOVA was performed on this 3*3*2*2 design (Appendix Table 5.12). It was expected that ambiguous terms would yield higher feature rates and higher feature frequencies in the shared set than unambiguous terms.

Table 5.12 (Appendix) allows the following observations:

(I) Ambiguous terms rendered more (weighted) features than nonambiguous terms for literal *A-terms*, metaphoric *A-* and *B-terms* in *single term* and *expression*, and for metaphoric *B-terms* in *context*. Ambiguous terms rendered fewer features than nonambiguous terms for literal *B-terms* in *single term* and *expression*, and in *context* for literal *A-* and *B-terms* and metaphoric *A-terms* (interaction of condition by ambiguity by term by expression type: $①_h$). (II) Ambiguous terms rendered more features than nonambiguous terms for metaphors and anomalies in *single term*. The same was valid for metaphors in *context*, whereas anomalies showed inverted effects (interaction of condition by ambiguity by expression type, parameter 8, liberal α-level: $②_h$). (III) Ambiguous terms rendered more features than nonambiguous terms for anomalous *A-* and *B-terms* and for literal *A-terms*. For literal *B-terms*, however, opposite effects were obtained (interaction of ambiguity by term by expression type, parameter 2: $③_h$). (IV) For literals and anomalies, ambiguous terms rendered higher frequencies of features in the shared set than nonambiguous terms. For metaphors, however, ambiguous terms rendered lower frequencies of features in the shared set than nonambiguous terms (interaction of ambiguity by expression type: $④_h$).

Lexical ambiguity did have effect on feature elicitation. However, the direction of the effect - stimulating or suppressing feature activation - was rather divergent, and therefore, may be considered a noise factor. If at all, metaphors benefited most from ambiguous terms in raising the number of features, while *context* had a disambiguating effect. Ambiguous terms contributed features with higher frequencies to the shared set than unambiguous ones, except for metaphors.

Check 5: Instance-category dominance

The last control concerns instance and category dominance of terms. It should be recalled that instance dominance (the number of times that the instance is mentioned, given the category) and category dominance (the number of times the category is mentioned, given the instance) has been found to speed up sentence verification time (Balota & Chumbley 1984). Since dominance has a pronounced effect on reaction time (RT), feature elicitation data might shed light on the extent to which a term is dominant.

The relation between instances and categories is different for the expression types. Literals match an instance as *A-term* with an appropriate higher-order category as *B-term* ('poetry is an art'). Metaphors match two inappropriate terms, either instances or categories ('poetry is a beast'). Anomalies match two inappropriate terms, of which the *A-term* is an instance of a higher-order category, whereas the *B-term* is an instance of a lower-order category ('poetry is a clack').

Since instance and category have a regular relation in literals, it can be expected that literal *A-* and *B-terms* are most associated. Metaphoric *A-* and *B-terms* are less associated, and thus, dominance will be less, whereas it will be least for anomalies. This pattern should show up in effects with expression type.

For each term, an index of dominance (associability) was established (Table 5.13) by counting the frequency of the *A-term* in response to the *B-term*, including relations ($A_{dom} = \sum A\varepsilon(Y \cup Y_r)$), and the frequency of the *B-term* in response to the *A-term*, including relations ($B_{dom} = \sum B\varepsilon(X \cup X_r)$).

Table 5.13: Grand means of *A-* and *B-term* dominance for all expression types and conditions.

Grand mean number of times that *A-* and *B-term* are mentioned in response of each other

	L		M		A	
	A_{dom}	B_{dom}	A_{dom}	B_{dom}	A_{dom}	B_{dom}
single term	.152	.096	.040	.033	.004	.000
expression	.231	.276	.714	.907	.452	.521
context	.341	.276	.352	.362	.352	.010

MANOVA in Table 5.14 (Appendix) suggests that:

(I) In *context*, *A-term* dominance was stronger than *B-term* dominance. At a liberal α-level, the same occurred in *single term*, and the reverse in *expression* (interaction of condition by term). (II) In *expression* (and *context* at liberal α), *A-* and *B-term* dominance was stronger in metaphors than in literals, whereas in *single term*, *A-* and *B-term* dominance was stronger in literals (interaction of condition by expression type). (III) *A-term* dominance was stronger than *B-term* dominance in literals and anomalies, whereas *B-term* dominance was stronger for metaphors (interaction of term by expression type, liberal α).

The results of instance-category dominance do not show a clear-cut picture. **The expression types in the present stimulus set do not correlate with specific dominance distributions. Often, the effects were too weak for a conservative α-level. Therefore, instance-category dominance was not a systematic effect, and may only add to the noise**.

In conclusion, the checks on the feature elicitation results suggested that (1) asymmetry hardly played a role in processing the present stimulus set, (2) repetition may have interfered with differences between conditions (*single* *term* vs *expression* and *context*), although repetitions within condition had no effect, (3) word frequency probably did not explain the low feature rates of anomalous *B-terms*, (4) lexical ambiguity and instance-category dominance had unsystematic effects. **With regard to the RT and EEG experiments, no systematic bias is expected from the potentially intervening variables.**

5.5 **General discussion for psychology**

In Section 5.3, three measures were outlined to determine the shared set. The first (M1) was based on strictly equal lines, the second (M2) compared words, considering derivations and declensions as equal features. M3 was a lenient measure, and compared letter clusters irrespective of meaning. It was found (Appendix Table 5.2) that M2 was most sensitive, suggesting that subjects usually conceived of features as words, rather than lines or letter clusters. Lines may contain too much undifferentiated information, whereas letter clusters may contain too little information. Moreover, a lenient definition of equality was less effective to calculate the shared set. Features may only be common when related in meaning.

The metaphor models were investigated with measure M2. **The comparison model was refuted**, because literal expressions did not have a larger shared set than metaphors, although both had larger shared sets than anomalies. Feature sets and shared sets were not fixed either, although it could be countered that fixed feature sets were activated, yet not all features were reported. The shared sets did not correspond with the interpretation of the expressions. More features of the shared set were reactivated in the interpretation of metaphors than of anomalies. However, the number of reactivated shared features was a small proportion of the total number of features in the interpretation.

The interaction model was strongly refuted, because relations were not created more frequently for metaphoric *B-terms* than for other term types, not even in *context*. Moreover, whatever combination of relation, ground or attributive sets were used, the number of matches never distinguished metaphors from other expression types. As a general finding, more relations were created as more context was added. Additionally, no kind of shared set - combining relations, grounds or attributive features - that was retrieved in the interpretations could distinguish expression types. As predicted by the interaction model, relations were created later in the associative flow. However, this was not typical for metaphors and occurred independent of condition, i.e. *context*.

The anomaly model received some support, when assuming two new shared sets, consisting of literal features for the one term, and figurative features for the other. Shared sets that consisted merely of figurative features did not differ among expression types. Shared sets that only contained literal features positively distinguished literal expressions and metaphors on the one hand from anomalies on the other. Metaphors had the largest shared sets of mixed literal and figurative features. Feature sets and shared sets were not fixed. Again, this may be due to a difference between activation and report.

With regard to the interpretations of the expressions, the anomaly model was the best among poor achievers. More features of the shared figurative-literal set were found in the interpretation of metaphors than of anomalies. More features of the shared literal-figurative set were found in the interpretation of metaphors than of literal expressions. However, the number of shared features that recurred in the interpretation was a small proportion of the total number of features in the interpretation.

At the initial serial positions of feature activation, more literal than figurative features were activated. The expected shift to figurative dominance at later positions was found only for anomalous *B-terms* in *expression* and *context*.

Note, however, that certain variables may have interfered with these results. Differences between *single term* on the one hand, and *expression* and *context* on the other might also have been due to repetition effects. Certain effects might have come out better, when noise factors such as word frequency, lexical ambiguity and instance-category dominance would have been totally excluded from the stimulus set. On the other hand, when effects are not robust, the models have little generalizing power.

A parallel two-stage anomaly model?

How can an anomaly model handle the problem that even for literal expressions, literal and figurative features were combined in the shared set, although the literal and figurative features are supposed to be activated in two serial stages? One solution may be that the literal and figurative stage operate in parallel. The serial two-stage anomaly model assumes that there are two information accumulators: The literal shared set and the figurative shared set, which are established in succession. However, it could also be that both accumulators work simultaneously.

In this case, literal and figurative features are activated in parallel in the *activation phase*. In the beginning, literal features may prevail. However, they are not all elicited before the activation of figurative features starts.

In the *comparison phase*, literal features may be compared with each other to calculate the literal shared set. At the same time, however, they are also compared with figurative features to determine the (mixed) figurative shared set.

Anomalies establish the smallest shared sets of either feature kind. They have small literal and small figurative feature sets, so that 'end of file' is reached soon while looking for shared features.

Literal expressions accumulate large literal shared sets, whereas the (mixed) figurative shared sets remain small. Literal expressions have large literal, but small figurative feature sets, so that 'end of file' is reached soon for the latter set type while looking for shared features.

Metaphors accumulate equally large literal shared sets. However, they are not recognized as literal expressions, because they accumulate large (mixed) figurative shared sets as well. Yet, literal and figurative set sizes for metaphors equal those for literal expressions, so that the processing speed of both expressions types may be the same. Thus, decisions between literal expressions and metaphors depend on the nature of the shared sets (literal vs literal **and** mixed figurative), whereas speed depends on the size of the feature sets.

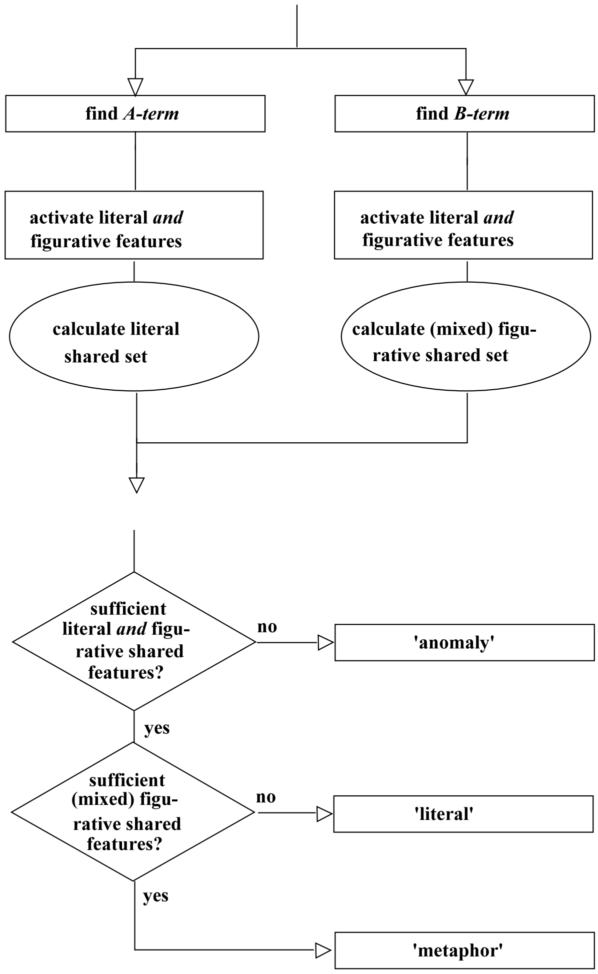
In the *response phase*, the criteria q_l and q_f may be equally checked in parallel. Literal expressions meet the literal criterion q_l , but do not meet q_f . The process stops, because meeting q_l is sufficient to qualify the expression as literal. Anomalies fail to exceed both criteria, and are dismissed as either literal or metaphoric. Metaphors satisfy both q_l and q_f . If so, **metaphors are not reconsidered anomalies** as the serial anomaly model presumes. However, they are expressions that exceed two decision criteria (Figure 5.19).

It may be, then, that anomalies are rapidly recognized, because both kinds of shared sets are small, quickly ascertained, checked and rejected as literal nor metaphoric. Literal expressions and metaphors may be equally fast, because the set sizes of literal and figurative features are equally large and 'end of file' is reached equally soon. Yet, the nature of the shared sets separates the decisions. In addition, it might be that metaphors are a little slower than literal expressions, if the accumulation of two large shared sets asks more effort than the accumulation of one.

Note that **the explanation of differences in RT and decisions by the parallel model is completely different from that by the serial model**. In the serial model, the differences derive only from the fixed order in which the response criteria are checked. In the parallel version, the speed of processing also depends on the size of the feature sets, and thus, on the time to accumulate shared features. The larger the sets and the more shared features are found, the slower the process.

Responses in *expression* may be faster than in *context*, because *context* is a more diffuse primer of the *B-term*. In *expression*, only the *A-term* is the prime. Chapter 6 will further investigate these possibilities.

Figure 5.19: Preliminary anomaly model for the parallel processing of two information sources: Literal versus (mixed) figurative shared features.



5.6 General discussion for the theory of literature

> (...) each metaphor can be traced back to a subjacent chain of metonymic connexions which constitute the framework of the code and upon which is based the constitution of any semantic field (...). (Eco 1979: 68; Culler 1981: 199-200)

> A metaphor can be invented because language, in its process of unlimited semiosis, constitutes a multidimensional network of metonymies (...). (...) the subjacent network of arbitrarily stipulated contiguities. (Eco 1979: 78; Culler 1981: 201)

Words evoke words, and in doing so, they create a set of related words. Culler stated that Eco perceived the connections between words (or items) as a measure of distance (or contiguity):

> Eco inverts the Jakobsonian relationship, because he thinks of systems and codes as spatial. If the system is spatial, then relations between items in the system may be thought of as relations of contiguity and hence as metonymic. (Culler 1981: 201)

In other words, Eco stipulated a distance function (the metonymy) between words in the language system. Since words supposedly have different contiguities, semantic equivalence or similarity between words is a function of their relative distances. How can such distances between words be estimated?

Distance metrics usually define a dimensional space in which similar stimuli occupy contiguous regions. These dimensions are defined by scaling the words (or other stimuli) for a number of oppositions (in the sense of Lotman). However, the choice of oppositions is free. Usually, they contain dimensions such as 'positive-negative', 'active-passive', and 'powerful-weak'.[10] Tourangeau & Sternberg (1982), for instance, used oppositions such as 'noble-ignoble' and 'warlike-peaceful' to scale the *A-* and *B-terms* of metaphors.

By means of what are termed 'factor scores', the general determinant in the scores of these dimensions is deduced. If the *A-term* is 'shark', it may score high on 'active', 'bad', and 'strength', whereas the *B-term* 'butterfly' would yield low scores for these poles. Thus, 'shark' and 'butterfly' are rather distant, and make a worse metaphor than e.g., 'seahorse' and 'butterfly'.

In other words, Eco's presumption that words in the system maintain certain distances can be operated by factor scores derived from preconceived dimensions, which form the coordinates in a dimensional space. The closer these coordinates are, the higher the semantic equivalence (or similarity) between the words.

Nevertheless, there is a problem in envisioning similarity as dimensional distances. Although stimuli may be closely related, they do not need to be similar or equivalent. Section 5.2 gave the example of 'death' and 'guilt', which are part of the same chain of metonymic connections, although they are not equivalent at all.

Eco's proposal circumvents this problem, if distance is replaced by the shared feature set. The feature sets generated by subjects in response to a word are arbitrary in principal, because language is subjective, as Eco stated. They are infinite, due to the process of unlimited semiosis; each feature may be a word that evokes features itself.

However, the arbitrariness is not so austere if the feature elicitation results of a word are averaged over subjects. Certain features are high-frequent, whereas others are unique. Thus, words activate features more closely connected to that word than to other words, which reduces the 'whimsical character' of the set. Moreover, subjects drain their associations within one minute, so that feature sets hardly ever exceed 15 lines of associations. Therefore, every feature set is demarcated by the knowledge of the speech community in question. Although in *la langue*, the 'chain of metonymic connexions' (the feature set) is arbitrary and infinite, in *la parole*, it is not.

The strength of the connections between word and feature can be expressed by the frequency that a feature occurs in response to the word. Since the features themselves are words capable of evoking features, each word can be activated as a feature of another word. Thus, the strength of the connections between words can be expressed by the frequency that a word occurs as a feature of another word. This mechanism is called 'response availability' or 'associability' and may also be reflected in the reaction time to recognize or produce a word given another word.

The strength of the connections among *words* can be established by calculating the overlap between the applicable feature sets. In Section 5.1, it was stressed that Jakobson's (1981: 27) ideas on syntagma and paradigm can be translated into terms of set theory. According to Culler (1981: 200), Jakobson connected metaphor with the paradigmatic axis of phonological, semantic and syntactic similarity. In other words, *A*- and *B-term* activate a set of features with high 'response availabilities' (the 'within paradigm' similarity). Metonyms are connected with the syntagmatic axis, and the possibilities of combining the terms to form sequences are increased when the set of shared features is large. Thus, the selection from the paradigmatic axis is based on finding sufficient features that match the features of a word in another paradigm. Thus, *selection* is directed towards high 'between paradigm' similarity in order to establish *combination*. Accordingly:

> Jakobson links metaphor more closely with the linguistic code, since relations of similarity occur primarily in the code or system. Only in poetic language are relations of similarity elements of the sequence. Metaphor is linked with *la langue* and metonymy with *la parole*, since relations of contiguity are manifested in the actual combinations of speech sequences. (Culler 1981: 200-201)

Eco also suggested that feature overlap between two 'relations of contiguity' constitutes the metaphor:

> Codes connect the notion of 'a beautiful woman' with the feature 'a long white neck'; they also connect 'swan' with 'a long white neck' and thus make possible through these two relations of contiguity, the metaphorical substitution of swan for woman. (Eco in Culler 1981: 201)

Thus, 'a long white neck' is the shared feature that connects 'swan' with 'woman' and comprises the metaphor. In Section 5.4, this idea initiated a full scale test of the metaphor models by means of a feature elicitation experiment. However, establishing meaning (the *signifié*) - either heuristically or computational - is based on the analysis of the form (the *signifiant*). To calculate the shared set, Section 5.3 defined three measures, based on different definitions of features and equality.

Measure M1 compared lines like 'a long white neck' for strict equality. Thus, letter string and position in the line should be equal, before the line was recognized as a feature shared between 'woman' and 'swan'. The second measure M2 was defined in compliance with Foucault (1988: 57), who hinted at a partial resemblance between elements, or 'similitude'. This measure was word-oriented, and recognized that 'white neck' was partly equal to 'whitened necks'. Shared features thus could formally differ in inflection or derivation, as long as their dictionary meaning was related. M3 was based on the formal-semantic equivalence theory of Jakobson, and compared letter clusters independently of meaning connectedness. In other words, 'white' was equal to 'whit', 'hit' and 'it', thus magnifying the shared set.

Although the Jakobsonian criterion found the most overlapping features, Table 5.2 shows that it was not the most sensitive measure. Since every formal equivalence was counted as a semantic equivalence (a matching feature), the differences between expression types faded. Thus, anomalies yielded as many shared features as literals and metaphors. The formal-semantic equivalence premise, thus blurred the differences among expression types with semantically unrelated shared features, although subjects indicated in Table 4.1 (Chapter 4) that these differences were relevant.

The measure derived from Foucault (M2) was a more sensitive measure. Subjects contrived of features as words, and less as lines or letter clusters. In other words, lines were probably split into more informational units, whereas letter clusters carried too little information to be of interest for a natural reading situation. Moreover, a strict formal equality (M1) was too severe, since subjects apparently combined words and their derivations or declensions as matching features.

How were metaphors distinguished from literals and anomalies, according to the results in Figure 5.0 up to 5.13? Culler stated that metaphor is hard to distinguish from literal expressions, because metaphor is an exemplification of creative cognitive processes.

> (...) metaphor [is located] in the gap between sense and reference, in the process of thinking of an object, event, or whatever, *as* something (...). (...) one can argue that cognition itself is essentially a process of seeing something as something. Metaphor thus becomes an instance of general cognitive processes at their most creative or speculative.

> However, precisely because this approach assimilates metaphor to general cognitive processes, it makes it difficult to establish any firm distinction between the literal and the metaphorical. (Culler 1981: 202)

Figures 5.8 and 5.9 indicate that a firm distinction between the literal and the metaphorical was found in the mixed set of shared literal and figurative features. Certain literal features of the *A-term* were figurative for the *B-term*, and certain literal *B-term* features were figurative for the *A-term*. These mixed shared sets were the largest for metaphors. Literal expressions had small mixed shared sets. Large sets were obtained for purely literal overlap and for the undifferentiated shared set defined by the comparison model. Anomalies had the smallest shared sets for all feature kinds.

Moreover, when metaphors were interpreted, it was found that the mixed shared sets were reactivated, and that they differentiated metaphors from other expression types (cf. Appendix Table 5.6). Since the mixed shared sets seem so important for distinguishing metaphors, a mechanism may be assumed that transformed literal features into figurative and vice versa from the one term to the other:

> In the case of the literal versus the figurative, the terms in which the figurative is defined so as to be distinguished from the literal lead one, paradoxically, to recognize the primacy of the figurative, either by identifying it with general cognitive processes and seeing the literal as figures whose figurality has been forgotten, or else by focusing on cases of catachresis where the figure seems to work without being contrasted with the literal. (Culler 1981: 206-207)

However, the reactivated shared sets were never equally large as the feature sets generated as interpretation. On the contrary, they were so small, that interpretation cannot be explained merely by finding the shared features. Culler is correct in asserting that:

> (...) there is no reason to claim that any particular similarities are part of the meaning of the sentence; the sentence says simply that there is similarity. In the case of a metaphor, however, (...) theorists usually insist that to give the meaning of the metaphor is to identify the similarities in question (...). (Culler 1981: 207)

Surprisingly, metaphors were not different for the purely figurative overlap. Nor did they stand out in creating more relations or forming more relation connections in the sense of M. Black. Other kinds of relation, ground or attributive connections did not yield the expected results either. Thus, expression type differences always reserved a role for literal or undifferentiated features. Since the mixed shared sets were prevalent in metaphors, Lotman's (1976: 100) postulate should be subscribed that 'this series of meanings' (the figurative) 'does not abolish linguistic meanings' (the literal) 'but co-exists with them forming a mutually correlated pair'.

In the experiment on judgements of similarity and figurativeness (Table 5.7), it was found that literal expressions were the highest in similarity, followed by metaphors and anomalies. As expected, metaphors had the highest figurativeness. Combined with the results of the feature elicitation experiment, the following picture emerged.

Metaphors had the largest mixed shared sets, which corresponded with a high figurativeness, whereas literal expressions were high in similarity. Since mixed overlap for the literal expressions was small, similarity probably corresponded with the large literal or undifferentiated overlap. Anomalies had small shared sets, regardless of feature type, thus receiving the lowest rates of similarity and figurativeness.

For that matter, Lotman's (1976: 100) statement should be reiterated that a poetic phenomenon such as metaphor activates linguistic associations (literal features), which are opposed to 'integrating' (figurative) and relational features. Lotman posited that the latter two feature types follow a 'different synthetic principle and are not analyzable into the mechanical sum of meaningful units'.

However, the results discussed above show that there probably is a correlation between the perceived similarity and figurativeness and the size of the assorted shared sets. Although the results were administered in a between-subject design, the mechanical sum of meaningful units may yet be established in a within-subject design. Subjects may elicit features for a sample of stimuli, and rate the similarity between the pairs. Semantic equivalence or similarity (*SIM*), then, is a function of the number of weighted shared features (W_S), which may be literal, figurative or relational: $SIM = \alpha\ W_S + \beta$, where α is the tangent or slope between the scores for *SIM* and W_S , and β is the interception from regression line to the origin. Thus, equivalence (or similarity) can be expressed by a 'mechanical sum of meaningful units'.

Another point of investigation was the potential asymmetric trait in metaphor processing. Tversky (1977) found that judgements of similarity are not necessarily symmetric. Sometimes, subjects judge that the *A-term* is more similar to the *B-term* than reversely. Tversky argued that asymmetry was predicated by the focus of attention. Subjects would prefer the unsalient (or feature-poor) *A-term* as the focus of comparison, and the salient (or feature-rich) as the frame of reference.

Since Tversky's findings, asymmetry has been investigated for metaphors and similes by, among others, Malgady & Johnson (1980) and Ortony, Vondruska, Foss & Jones (1985). In the present study, however, no asymmetry was found in judgements of similarity and figurativeness. The judgements from *A*- to *B-term* were not reliably different from *B*- to *A-term* judgements. The general discussion for psychology suggests that the studies that found asymmetry used biasing stimuli, and sometimes performed a weaker form of statistics.

In conclusion, **the comparison model was repudiated**, because the distinction in shared set size between literals and metaphors failed. Despite the finding that anomalies had the smallest shared set, literals and metaphors did not differ. The feature sets were not fixed. However, it could be said that all features may be activated, yet not written down by the subjects. The features in the interpretation of the expressions were not identical to the shared set. When the shared sets were used to search the interpretation sets, only one expression type difference occurred. More shared features were present in the interpretation of metaphors than in that of anomalies.

The interaction model was even less successful, because metaphoric *B-terms* did not stimulate the creation of relations more than the other expression types, not even in *context*. As a general finding for all expression types, more relations were created when more context was added. However, the number of matches between all possible combinations of relation, ground or attributive sets never differed among expression types. Moreover, these matches were not equal to the interpretation of the expressions, and those that were retrieved made no distinctions among expression types. Relations were created at later positions in the associative flow, as expected by the interaction model. However, this was not typical for metaphors and occurred independent of condition, i.e. *context*.

The anomaly model was partly affirmed, however, only because a new shared set was defined of features that were literal for the one term and figurative for the other. Purely figurative shared sets could not discriminate the expression types. Purely literal shared sets failed to distinguish literal expressions from metaphors. Both had large literal overlap, larger than for anomalies. Only the shared sets of mixed literal and figurative features could detect metaphors, which were larger than for the other expression types. Again, feature sets and overlap were not fixed. As indicated earlier, however, activated features may not have been reported by the subjects.

The shared sets did not correspond with the interpretation. Compared with anomalies, more features of the shared figurative-literal set of metaphors were found in the interpretation. Compared with literal expressions, more features of the shared literal-figurative set of metaphors were found in the interpretation.

Literal features were dominant over figurative features at the first positions in the feature list. Figurative features accumulated at later positions and dominated literal features only for anomalous *B-terms* in *expression* and *context*.

Apparently, the anomaly model is the best among poor performers. Moreover, when the serial anomaly model is changed into a parallel model, it can account for most of the feature elicitation results.

In the *activation phase*, literal and figurative features are activated in parallel by *A*- and *B-term*. At first, literal features arise in the associative flow, while figurative features rise to their peak somewhat later.

In the *comparison phase*, literal expressions and metaphors find equally large literal overlap. Yet, metaphors are not categorized as literal expressions. This can only be explained when a parallel comparison for literal and figurative overlap is conjectured. In this case, literal expressions, metaphors and anomalies establish literal as well as figurative overlap. Literal expressions find many shared literal features and few shared figurative features. Anomalies find the least shared features of either kind. Metaphors, however, find literal as well as figurative overlap. According to the results, the figurative overlap is composed of features that are literal for the one term and figurative for the other. In other words, the decision between literal expressions and metaphors depends on the nature of the shared sets (literal vs literal **and** mixed figurative).

Processing speed depends on the size of the feature sets. Anomalies have small sets, so that an exhaustive search for shared features ends quickly. Literal expressions and metaphors have equally large literal and figurative sets, so that the search for shared features takes equally long. Metaphors may be understood slightly slower if the accumulation of two large shared set requires more effort than the accumulation of one.

In the *response phase*, two threshold criteria are at work. The criterion for literal overlap q_l and for figurative overlap q_f. Literal expressions satisfy the literal criterion q_l easily, but q_f is not reached. Yet, meeting q_l is sufficient to be judged as literal, so that the process stops. Anomalies fail to exceed both criteria, and terminate all further processing. Therefore, **metaphors are not reconsidered anomalies**, but expressions that exceed two decision criteria.

Moreover, context may influence the process. Ricœur's (1974: 104) view of the role of contextual c(l)ues was quite apt. His statement purported that 'a clue is a kind of index for a specific construction, both a set of permissions and a set of prohibitions; it excludes some unfitting constructions and allows some others that make more sense of the same words'.

This is illustrated by the three conditions in the experiment. In *single term*, personal associations could be generated freely; however, this condition also prohibited ample semantic connections. Many features were activated, whereas only small shared sets were found. *Expression* prohibited the free elicitation of features, because the terms settled the limits within which the activated features were meaningful. However, since merely appropriate features were activated, many of them intersected. Both terms evoked a small number of mainly identical features, so that the shared sets were large. *Context* permitted the activation of many associations in response to the terms and the clues; however, the clues also primed the semantic connections, so that a large overlap was found.

In other words, if the parallel anomaly model holds, the RT experiments should show that anomalies are the most quickly identified, followed by literals and metaphors. Moreover, the priming of the *B-term* in *expression* is less cryptic than in *context*. In *context*, 'ambiguity is an intrinsic, inalienable character of any self-focused message, briefly, a corollary feature of poetry' (Jakobson 1981: 42). In *expression*, only the *A-term* primes the *B-term*, whereas in *context*, the priming is by many (ambiguous) words. Thus, decisions in *expression* may be quicker than in *context*.

Regardless of all this, Ricœur (1974: 104) was again right when he attested that 'a literary work is more than a linear succession of sentences. It is a cumulative, holistic process'. None of the theories can explain interpretation, because none of the shared sets formed a major part of the interpretation feature set. Thus, finding shared features or perceiving similarity is inadequate as an explanation for the interpretation of expressions. Hoorn & Veldman (in preparation) studied the compounds of interpretation sets, to ascertain the percentage of shared, distinctive and newly activated features in the interpretation.

Notes:

1. It could be objected that salience is attributive to intensity, so that a loud tone is more salient than a soft tone, thus making the argument unfeasible that familiarity or communality are good alternatives. Although intensity may well be an indicator of salience in the case of sound and light, for words - i.e. feature elicitation data - intensity is hard to make operative, because it is not a factor that can be additively increased as sound or light. A word cannot be made 'more intense' by turning a control knob.

2. Gleitman, Gleitman, Miller & Ostrin (1996) showed (experiment 5) that asymmetry may be found while using comparisons with nonsense words ('the ZUM is identical to the GAX'), thus demonstrating that even the position of the noun in the sentence may contribute to asymmetry perception.

3. In the original paper, a minus sign is in front of $U(A)$, which must be a misprint.

4. Cf. 'The green sea is my girl friend'. Charles, J.B. (1987). Een suite van de zee. *De groene zee is mijn vriendin, gedichten 1944-1982*, De Bezige Bij, Amsterdam.

5. It may be argued that a feature is not added by an adjective, but that it is selected from the set of the noun, thus making the feature more salient.

6. Steen (personal communication) objected to the idea that the relations shown in Table 5.1 are identical to the relations that M. Black drove at. They should be analogies. However, subjects were instructed beforehand with an analogy used by interaction theory (Tourangeau & Sternberg 1982), saying that teachers have power over their classes as kings have over their empires. They were urged to make relations similar to 'power' among features and terms. After that, it is not for the researcher to decide whether these are proper instantiations of Black's theory or not, unless theory provides a characteristic that distinguishes an analogical relation from any other relation.

7. An objection may be that Clark assumes the variance-covariance matrix to be equal. He uses univariate tests, which demand a compound symmetry (equal variances and equal correlations). Data that meet this demand are empirically almost nonexistent. Therefore, it should be recommended that multivariate tests with more variance components become available for standard statistical packages on PC.

8. Steen (personal communication) emphasized that there may be an interfering effect of time pressure during the creation of relations. Since the production of relations is an effortful procedure, subjects might have created fewer relations than under more natural circumstances. This may explain the absence of differences among expression types for relation set size and shared relation set size.

It could be countered, however, that subjects had no time pressure in creating relations. After feature elicitation (one minute), they were allowed to make relations at their own speed, exactly to avoid this 'ceiling effect' for the harder task.

9. *Interpretation* was designed as a within-subject factor level, because it is implausible that the shared features found for subject 1 will be retrieved in the interpretation of subject 2. A within-subject design will stimulate the size of $X \cap Y \cap Z$. Yet, the size of this shared set was so small, that interpretation could not be explained by it, not even in a within-subject design.

10. Following in the wake of Osgood, Suci & Tannenbaum (1957), such a bipolar scale is also called a semantic differential e.g., in Hauptmeier, H. & Schmidt, S. J. (1985). *Einführung in die Empirische Literaturwissenschaft*. Vieweg & Sohn, Braunschweig, Wiesbaden, 154.

Appendix to Chapter 5

Table 5.2: F-tests on the effects of condition, expression type and term on the grand means of weighted feature set size (W_X and W_Y) and weighted shared set size (W_S), using the measures M1, M2, M3. Features (*x*, *y*) could be literal (*l*) or figurative (*f*), and could be followed by relations (*r*). Coeff. = the standardized discriminant function coefficients, corr. = correlations between M1, M2, M3 and discriminant functions. Coefficients and correlations are tabulated only for the significant effects. Measures with the highest distinctive power for an effect are labeled with an asterisk.

Main effects of condition: *Single term* vs *expression* vs *context*

dep. variable	Pillai's Trace	$F_{6,72}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X+W_Y	.75	7.29	.000	.66	-.90	- .88	-.96 *	- .75	-.99
$W_{Xl}+W_{Yl}$	.56	4.77	.000	7.41	.59 *	-7.46	.50	.61	.49
$W_{Xf}+W_{Yf}$	.56	4.69	.000	8.90	-.25	-12.59	-.35 *	3.42	-.35
$W_{Xr}+W_{Yr}$	.41	3.10	.009	1.15	-.56	-6.44	-.74 *	4.62	-.68
W_X+W_{Yr}	.67	6.16	.000	5.31	.23	-6.37	.05 *	1.25	.06
W_S	.58	5.02	.000	1.23	-.75	-2.48	-.92 *	.41	-.88
W_{Sll}	.38	2.86	.015	1.62	-.58	-3.85	-.77 *	1.64	-.63
W_{Sff}	.35	2.60	.024	1.64	-.70	.80	-.83	-3.25	-.87 *
W_{Srr}	.51	4.12	.001	- .06	-.67	4.83	-.88	-5.69	-.91 *
W_{Sxr}	.76	7.46	.000	- .13	-.77	1.16	-.94	-2.02	-.98 *

Interactions of condition by expression type:

dep. variable	Pillai's Trace	$F_{12,66}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X+W_Y	.80	3.70	.000	-1.37	-.23	4.62	-.22 *	-3.73	-.48
				- .03	.02	.69	-.13	- .21	-.02
$W_{Xl}+W_{Yl}$	.39	1.33	.221						
$W_{Xf}+W_{Yf}$	.31	1.02	.436						
$W_{Xr}+W_{Yr}$	.72	3.09	.002	-3.13	-.35 *	8.87	-.08	-5.68	-.08
				2.15	.11	-7.88	.05	5.68	.05
W_X+W_{Yr}	.65	2.67	.006	3.46	.54	-4.82	.42 *	2.41	.44
				.89	-.26	-1.49	-.21	- .10	-.18
W_S	.47	1.73	.079						
W_{Sll}	.26	.83	.619						
W_{Sff}	.39	1.34	.213						
W_{Srr}	.52	1.93	.046	- .11	-.22	11.73	.17	-11.17	.14
				- .42	-.25	-10.34	-.25 *	10.03	-.21
W_{Sxr}	.40	1.40	.188						

Main effects of expression type: Literals vs metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{6,32}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X+W_Y	.69	12.38	.000	-2.00	-.63	1.96	-.60 *	- .80	-.74
				- .15	-.33	-1.69	-.50	1.64	-.36
W_{Xl}+W_{Yl}	.60	8.15	.000	-3.01	-.68	2.17	-.63	.44	-.66
				.71	-.39	-5.60	-.51 *	4.67	-.42
W_{Xf}+W_{Yf}	.14	.89	.510						
W_{Xr}+W_{Yr}	.37	3.24	.013	-1.50	-.00	13.66	-.01	-11.85	-.04
				1.58	-.26	-7.97	-.29 *	5.53	-.28
W_X+W_{Yr}	.35	2.94	.021	-2.61	-.66	4.67	-.56 *	-2.99	-.67
				- .29	-.12	- .28	-.35	.80	-.33
W_S	.66	10.61	.000	.21	-.36	1.05	-.34	-2.05	-.42
				1.84	-.36	-3.85	-.48 *	1.32	-.47
W_{Sll}	.59	7.78	.000	-1.17	-.61	2.79	-.58	-2.54	-.69 *
				-1.00	-.30	1.48	-.34	-1.16	-.31
W_{Sff}	.20	1.33	.270						
W_{Srr}	.31	2.48	.044	1.06	.22	13.23	.28 *	-13.02	.25
				- .97	-.20	-5.71	-.33	5.34	-.32
W_{Sxr}	.35	2.89	.023	1.60	.41	8.30	.22	-9.11	.18 *
				- .83	-.11	.74	-.23	- .82	-.25

Interactions of condition by term:

dep. variable	Pillai's Trace	$F_{6,72}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X - W_Y	.74	7.06	.000	.62	.47	-1.85	.48	1.95	.82 *
W_{Xl} - W_{Yl}	.12	.79	.574						
W_{Xf} - W_{Yf}	.14	.90	.494						
W_{Xr} - W_{Yr}	.24	1.64	.147						
W_X - W_{Yr}	.55	4.63	.000	1.43	.98 *	- .83	.90	.40	.83
W_S	.29	2.10	.063						
W_{Sll}	.21	1.41	.220						
W_{Sff}	.29	2.05	.070						
W_{Srr}	.16	1.07	.385						
W_{Sxr}	.39	2.91	.013	- .61	-.94 *	- .09	-.91	- .38	-.85

Main effects of term: *A-term* vs *B-term*

dep. variable	Pillai's Trace	$F_{3,35}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X - W_Y	.78	41.54	.000	.47	-.93	-1.79	-.98 *	.39	-.80
W_{Xl} - W_{Yl}	.37	6.86	.001	2.26	.96 *	- .65	.92	- .68	.84
W_{Xf} - W_{Yf}	.41	8.14	.000	1.79	.98 *	- .49	.94	- .33	.91
W_{Xr} - W_{Yr}	.15	2.19	.106						
W_X - W_{Yr}	.96	290.98	.000	-1.08	-.93	1.20	-.91	-1.15	-.94 *
W_S	.13	1.79	.166						
W_{Sll}	.01	.12	.946						
W_{Sff}	.19	2.78	.055	1.03	.25	-1.05	-.50 *	- .33	-.59
W_{Srr}	.10	1.31	.284						
W_{Sxr}	.53	13.16	.000	- .62	-.96 *	- .20	-.92	- .25	-.81

Interactions of condition by term by expression type:

dep. variable	Pillai's Trace	$F_{12,66}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X - W_Y	1.10	6.72	.000	- .26	.27	2.01	.26 *	-1.98	-.04
				- .03	.25	-1.02	.24	1.62	.44
W_{Xl} - W_{Yl}	.27	.87	.571						
W_{Xf} - W_{Yf}	.34	1.16	.328						
W_{Xr} - W_{Yr}	.48	1.77	.071						
W_X - W_{Yr}	.37	1.28	.251						
W_S	.44	1.57	.121						
W_{Sll}	.35	1.18	.309						
W_{Sff}	.40	1.39	.189						
W_{Srr}	.42	1.50	.147						
W_{Sxr}	.50	1.86	.055	2.26	.02	-2.60	-.31 *	- .09	-.27
				.84	.15	-1.03	.14	.50	.22

Interactions of term by expression type:

dep. variable	Pillai's Trace	$F_{6,32}$	p	M1 coeff.	M1 corr.	M2 coeff.	M2 corr.	M3 coeff.	M3 corr.
W_X - W_Y	.85	31.72	.000	- .74	-.30	1.81	-.24	-1.39	-.48
				- .92	-.69 *	- .75	-.68	1.28	-.48
W_{Xl} - W_{Yl}	.79	20.14	.000	- .03	-.39	1.52	-.35	-1.22	-.44
				-1.60	-.85 *	- .90	-.82	1.55	-.71
W_{Xf} - W_{Yf}	.25	1.78	.133						
W_{Xr} - W_{Yr}	.26	1.90	.110						
W_X - W_{Yr}	.35	2.90	.023	.68	.96	.73	.92 *	.44	.86
				- .37	-.16	.35	-.10	- .17	-.07
W_S	.14	.90	.507						
W_{Sll}	.10	.61	.716						
W_{Sff}	.17	1.14	.360						
W_{Srr}	.13	.86	.531						
W_{Sxr}	.25	1.78	.134						

Table 5.3: MANOVAs for the effects of condition, expression type and term on the grand mean weighted feature set sizes (W_X , W_Y ; W_{Xl} , W_{Yl} ; W_{Xf} , W_{Yf} ; W_{Xr} , W_{Yr} ; W_{Xg} , W_{Yg} ; W_{Xa} , W_{Ya}). Only the significant parameters are tabulated. Figures in black circles followed by subscript letters (for example, ❶$_a$) refer to important effects discussed in the text.

Main effects of condition:

1 = *single term* vs *expression*
2 = *single term* vs *context*
3 = *expression* vs *context*

dep. variable	$F_{2,37}$	p	df	quasi-F	p		coeff.	t	p
$W_X + W_Y$ ❸$_a$	3.70	.034	$_{2,37}$	3.70	.045	1	13.73	2.70	.010**
$W_{Xl} + W_{Yl}$ ❷$_a$	4.63	.016	$_{2,38}$	4.54	.020	1	15.77	2.60	.013**
$W_{Xf} + W_{Yf}$	2.07	.141							
$W_{Xr} + W_{Yr}$ ❺$_a$	6.12	.005	$_{2,40}$	5.91	.008	2	-9.85	-3.35	.002**
$W_{Xg} + W_{Yg}$ ❻$_a$	5.12	.011	$_{2,39}$	4.99	.009	2	-7.18	-2.89	.006**
$W_{Xa} + W_{Ya}$	11.94	.000	$_{2,37}$	11.80	.000	1	19.31	4.59	.000**

Interactions of condition by term:

1 = (*single term* vs *expression*) vs (*A*- vs *B-term*)
2 = (*single term* vs *context*) vs (*A*- vs *B-term*)
3 = (*expression* vs *context*) vs (*A*- vs *B-term*)

dep. variable	$F_{2,37}$	p	df	quasi-F	p		coeff.	t	p
$W_X - W_Y$	9.73	.000	$_{2,26}$	7.00	.005	1	3.68	4.40	.000**
						3	-2.57	-2.67	.011**
$W_{Xl} - W_{Yl}$	1.75	.188							
$W_{Xf} - W_{Yf}$	1.73	.191							
$W_{Xr} - W_{Yr}$	1.64	.208							
$W_{Xg} - W_{Yg}$	.13	.882							
$W_{Xa} - W_{Ya}$	8.39	.001	$_{2,40}$	7.68	.006	1	3.78	4.09	.000**

Main effects of term:

1 = *A*- vs *B-term*

dep. variable	$F_{1,37}$	p	df	quasi-F	p		coeff.	t	p
$W_X - W_Y$ ❾$_a$	126.87	.000	$_{1,10}$	74,26	.000	1	4.05	11.26	.000*
$W_{Xl} - W_{Yl}$	18.63	.000	$_{1,7}$	8.90	.024	1	2.11	4.31	.000*
$W_{Xf} - W_{Yf}$ ❹$_a$	23.15	.000	$_{1,20}$	16.95	.000	1	1.94	4.81	.000*
$W_{Xr} - W_{Yr}$	.94	.339							
$W_{Xg} - W_{Yg}$	.02	.888							
$W_{Xa} - W_{Ya}$	102.64	.000	$_{1,14}$	67.46	.000	1	4.03	10.13	.000*

Interactions of condition by expression type:

1 = (*single term* vs *expression*) vs (literals vs metaphors)
2 = (*single term* vs *context*) vs (literals vs metaphors)
3 = (*expression* vs *context*) vs (literals vs metaphors)
4 = (*single term* vs *expression*) vs (literals vs anomalies)
5 = (*single term* vs *context*) vs (literals vs anomalies)
6 = (*expression* vs *context*) vs (literals vs anomalies)
7 = (*single term* vs *expression*) vs (metaphors vs anomalies)
8 = (*single term* vs *context*) vs (metaphors vs anomalies)
9 = (*expression* vs *context*) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{4,74}$	p	df	quasi-F	p		coeff.	t	p
W_X+W_Y	.36	4.18	.004	$_{4,79}$	4.39	.005	2	-6.19	-3.92	.000***
							3	-6.00	-3.29	.002***
							5	-5.92	-3.00	.005***
$W_{Xl}+W_{Yl}$	.22	2.28	.068							
$W_{Xf}+W_{Yf}$	.08	.81	.519							
$W_{Xr}+W_{Yr}$	.04	.42	.789							
$W_{Xg}+W_{Yg}$	.15	1.49	.214							
$W_{Xa}+W_{Ya}$	.22	2.29	.067							

Main effects of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable		Pillai's Trace	$F_{2,36}$	p	df	quasi-F	p		coeff.	t	p
W_X+W_Y	➓ a	.56	23.66	.000	$_{2,74}$	24.70	.000	2	4.72	5.57	.000**
								3	5.00	6.79	.000**
$W_{Xl}+W_{Yl}$		.43	13.89	.000	$_{2,76}$	17.32	.000	2	3.84	4.75	.000**
								3	4.48	5.07	.000**
$W_{Xf}+W_{Yf}$		.03	.56	.571							
$W_{Xr}+W_{Yr}$		.05	.99	.380							
$W_{Xg}+W_{Yg}$		.05	1.01	.373							
$W_{Xa}+W_{Ya}$		.46	15.93	.000	$_{2,76}$	15.62	.000	2	4.67	5.55	.000**
								3	4.22	4.37	.000**

Interactions of condition by term by expression type:

1 = (*single term* vs *expression*) vs (*A*- vs *B-term*) vs (literals vs metaphors)
2 = (*single term* vs *context*) vs (*A*- vs *B-term*) vs (literals vs metaphors)
3 = (*expression* vs *context*) vs (*A*- vs *B-term*) vs (literals vs metaphors)
4 = (*single term* vs *expression*) vs (*A*- vs *B-term*) vs (literals vs anomalies)
5 = (*single term* vs *context*) vs (*A*- vs *B-term*) vs (literals vs anomalies)
6 = (*expression* vs *context*) vs (*A*- vs *B-term*) vs (literals vs anomalies)
7 = (*single term* vs *expression*) vs (*A*- vs *B-term*) vs (metaphors vs anomalies)
8 = (*single term* vs *context*) vs (*A*- vs *B-term*) vs (metaphors vs anomalies)
9 = (*expression* vs *context*) vs (*A*- vs *B-term*) vs (metaphors vs anomalies)

dep. variable		Pillai's Trace	$F_{4,74}$	p	df	quasi-F	p		coeff.	t	p
W_X - W_Y	❶a	.45	5.46	.000	4,72	7.47	.000	4	-1.92	-3.57	.001***
								7	-2.68	-4.15	.000***
								8	-2.70	-4.17	.000***
W_{Xl} - W_{Yl}		.14	1.45	.225							
W_{Xf} - W_{Yf}		.12	1.18	.322							
W_{Xr} - W_{Yr}		.14	1.43	.230							
W_{Xg} - W_{Yg}		.18	1.88	.123							
W_{Xa} -W_{Ya}	❼a	.37	4.27	.004	4,47	4.72	.005	4	-2.06	-3.98	.000***
								7	-2.61	-3.70	.001***
								8	-2.20	-3.11	.004***

Interactions of term by expression type:

1 = (*A*- vs *B-term*) vs (literals vs metaphors)
2 = (*A*- vs *B-term*) vs (literals vs anomalies)
3 = (*A*- vs *B-term*) vs (metaphors vs anomalies)

dep. variable		Pillai's Trace	$F_{2,36}$	p	df	quasi-F	p		coeff.	t	p
W_X - W_Y		.73	49.86	.000	2,19	33.93	.000	1	1.73	8.71	.000**
								2	-.83	-3.60	.001**
								3	-2.56	-9.21	.000**
W_{Xl} - W_{Yl}	❸a	.73	50.02	.000	2,10	20.20	.000	1	1.78	9.71	.000**
								2	-1.16	-4.24	.000**
								3	-2.95	-8.40	.000**
W_{Xf} - W_{Yf}		.04	.94	.399							
W_{Xr} - W_{Yr}		.10	2.17	.128							
W_{Xg} - W_{Yg}		.14	3.08	.058							
W_{Xa} - W_{Ya}		.69	41.85	.000	2,29	40.67	.000	1	1.60	7.56	.000**
								2	-1.15	-5.17	.000**
								3	-2.76	-9.07	.000**

Table 5.4: MANOVAs for the effects of condition, expression type and term on the grand mean weighted shared set sizes (W_S, W_{Sll}, W_{Slf}, W_{Sfl}, W_{Sff}, W_{Srr}, W_{Sxr}, W_{Sgr}). Only the significant parameters are tabulated. Black circled figures followed by a subscript (e.g., ❶$_b$) refer to important effects discussed in the text.

Main effects of condition:

1 = *single term* vs *expression*
2 = *single term* vs *context*
3 = *expression* vs *context*

dep. variable	$F_{2,37}$	p	df	quasi-F	p		coeff.	t	p
W_S ❶$_b$	22.61	.000	$_{2,17}$	12.48	.000	1	-3.02	-6.14	.000**
						2	-2.27	-4.61	.000**
W_{Sll} ❹$_b$	6.86	.003	$_{2,31}$	5.16	.011	1	-1.08	-3.39	.002**
						2	-.81	-2.53	.016**
W_{Slf} ❻$_b$	15.27	.000	$_{2,35}$	12.30	.000	1	-.38	-4.33	.000**
						2	-.41	-4.67	.000**
W_{Sfl} ❾$_b$	13.19	.000	$_{2,41}$	12.10	.000	1	-.47	-3.99	.000**
						2	-.51	-4.37	.000**
W_{Sff} ❷$_c$	6.83	.003	$_{2,20}$	4.17	.030	1	-.83	-3.09	.004**
						2	-.79	-2.93	.006**
W_{Srr} ❸$_c$	10.27	.000	$_{2,41}$	9.40	.000	1	-1.10	-2.66	.011**
						2	-1.79	-4.34	.000**
W_{Sxr}	26.69	.000	$_{2,27}$	18.59	.000	1	-1.54	-4.46	.000**
						2	-2.39	-6.94	.000**
W_{Sgr}	11.86	.000	$_{2,40}$	10.28	.000	1	-.69	-3.38	.002**
						2	-.91	-4.42	.000**

Interactions of condition by term:

1 = (*single term* vs *expression*) vs (*A*- vs *B-term*)
2 = (*single term* vs *context*) vs (*A*- vs *B-term*)
3 = (*expression* vs *context*) vs (*A*- vs *B-term*)

dep. variable	$F_{2,37}$	p	df	quasi-F	p		coeff.	t	p
W_S ❸$_b$	3.33	.047	$_{3,37}$	4.60	.009	3	-.06	-2.57	.014**
W_{Sll}	1.71	.195							
W_{Slf}	1.95	.156							
W_{Sfl}	.61	.547							
W_{Sff}	4.79	.014	$_{74,37}$	.50	>.50				
W_{Srr}	1.29	.288							
W_{Sxr} ❹$_c$	9.16	.001	$_{2,37}$	9.26	.000	1	.13	3.09	.004**
						2	.17	3.81	.000**
W_{Sgr} ❺$_c$	5.80	.006	$_{2,37}$	5.59	.010	1	.07	2.66	.011**
						2	.08	2.88	.006**

Main effects of term:

1 = *A*- vs *B-term*

dep. variable	$F_{1,37}$	p	df	quasi-F	p		coeff.	t	p
W_S ❻c	1.28	.265							
W_{Sll}	.11	.739							
W_{Slf}	.33	.570							
W_{Sfl}	.65	.424							
W_{Sff}	2.21	.146							
W_{Srr}	2.63	.113							
W_{Sxr}	35.87	.000	$_{1,37}$	35.90	.000	1	-.11	-5.98	.000*
W_{Sgr}	18.50	.000	$_{1,37}$	17.36	.000	1	-.05	-4.30	.000*

Interactions of condition by expression type:

1 = (*single term* vs *expression*) vs (literals vs metaphors)
2 = (*single term* vs *context*) vs (literals vs metaphors)
3 = (*expression* vs *context*) vs (literals vs metaphors)
4 = (*single term* vs *expression*) vs (literals vs anomalies)
5 = (*single term* vs *context*) vs (literals vs anomalies)
6 = (*expression* vs *context*) vs (literals vs anomalies)
7 = (*single term* vs *expression*) vs (metaphors vs anomalies)
8 = (*single term* vs *context*) vs (metaphors vs anomalies)
9 = (*expression* vs *context*) vs (metaphors vs anomalies)

dep. variable		Pillai's Trace	$F_{4,74}$	p	df	quasi-F	p		coeff.	t	p
W_S		.20	2.06	.094							
W_{Sll}		.11	1.10	.362							
W_{Slf}	❽b	.32	3.56	.010	$_{5,46}$	3.20	.019	1	.27	3.70	.001***
W_{Sfl}	❶c	.33	3.65	.009	$_{4,68}$	5.48	.001	1	.39	4.15	.000***
								7	-.31	-3.60	.001***
W_{Sff}		.18	1.87	.124							
W_{Srr}		.11	1.13	.345							
W_{Sxr}		.06	.61	.656							
W_{Sgr}		.08	.78	.541							

Main effects of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable		Pillai's Trace	$F_{2,36}$	p	df	quasi-F	p		coeff.	t	p
W_S	❷$_b$	.48	16.84	.000	$_{2,16}$	7.26	.007	2	.48	2.97	.005**
								3	.98	5.61	.000**
W_{Sll}	❺$_b$	.48	16.77	.000	$_{2,26}$	6.85	.009	2	.40	4.29	.000**
								3	.50	4.47	.000**
W_{Slf}	❼$_b$	.33	9.11	.001	$_{2,14}$	5.35	.021	1	-.13	-4.19	.000**
								3	.11	3.36	.002**
W_{Sfl}	❿$_b$	.47	16.23	.000	$_{2,47}$	17,76	.000	1	-.18	-4.65	.000**
								3	.21	5.77	.000**
W_{Sff}		.08	1.70	.196							
W_{Srr}		.14	3.13	.055							
W_{Sxr}		.07	1.42	.253							
W_{Sgr}		.08	1.71	.194							

Interactions of condition by term by expression type:

1 = (*single term* vs *expression*) vs (*A*- vs *B-term*) vs (literals vs metaphors)
2 = (*single term* vs *context*) vs (*A*- vs *B-term*) vs (literals vs metaphors)
3 = (*expression* vs *context*) vs (*A*- vs *B-term*) vs (literals vs metaphors)
4 = (*single term* vs *expression*) vs (*A*- vs *B-term*) vs (literals vs anomalies)
5 = (*single term* vs *context*) vs (*A*- vs *B-term*) vs (literals vs anomalies)
6 = (*expression* vs *context*) vs (*A*- vs *B-term*) vs (literals vs anomalies)
7 = (*single term* vs *expression*) vs (*A*- vs *B-term*) vs (metaphors vs anomalies)
8 = (*single term* vs *context*) vs (*A*- vs *B-term*) vs (metaphors vs anomalies)
9 = (*expression* vs *context*) vs (*A*- vs *B-term*) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{4,74}$	p	df	quasi-F	p
W_S	.28	3.04	.022	$_{9,18}$	2.00	.100
W_{Sll}	.22	2.32	.064			
W_{Slf}	.18	1.85	.127			
W_{Sfl}	.13	1.36	.253			
W_{Sff}	.21	2.20	.077			
W_{Srr}	.25	2.74	.034	$_{6,17}$	1.44	.291
W_{Sxr}	.10	1.01	.406			
W_{Sgr}	.10	.98	.420			

Interactions of term by expression type:

1 = (*A*- vs *B-term*) vs (literals vs metaphors)
2 = (*A*- vs *B-term*) vs (literals vs anomalies)
3 = (*A*- vs *B-term*) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{2,36}$	p
W_S	.02	.55	.581
W_{Sll}	.07	1.49	.239
W_{Slf}	.10	2.04	.144
W_{Sfl}	.12	2.58	.090
W_{Sff}	.11	2.35	.109
W_{Srr}	.07	1.41	.255
W_{Sxr}	.03	.57	.570
W_{Sgr}	.05	1.10	.343

Table 5.6: MANOVAs for the grand means of the weighted feature set size W_Z (= W_{Zl}+W_{Zf}), W_{Zr} and W_{Zg} , and of W_{T1} , W_{T2} , ${}_{Zl}W_{T2ll}$, ${}_{Zf}W_{T2ll}$, ${}_{Zl}W_{T2lf}$, ${}_{Zf}W_{T2lf}$, ${}_{Zl}W_{T2fl}$, ${}_{Zf}W_{T2fl}$, ${}_{Zl}W_{T2ff}$, ${}_{Zf}W_{T2ff}$, ${}_{Zr}W_{T2rr}$, ${}_{Zg}W_{T2rr}$, ${}_{Zr}W_{T2xr}$, ${}_{Zg}W_{T2xr}$, ${}_{Zr}W_{T2gr}$, ${}_{Zg}W_{T2gr}$. Except for W_Z , only the significant effects are exposed. Black circled figures followed by a subscript letter (e.g., ❶$_d$) direct to important effects discussed in the text.

Main effects of condition:

dep. variable		$F_{1,18}$	p	df	quasi-F	p	coeff.	t	p
W_Z		3.27	.087						
$W_{Zr}+W_{Zg}$	❿$_d$	5.73	.028	$_{1,18}$	5.69	.029	-3.81	-2.39	.028*
${}_{Zf}W_{T2ll}+W_{Zf}$		10.01	.005	$_{1,19}$	9.24	.006	-2.44	-3.16	.005*
${}_{Zf}W_{T2lf}+W_{Zf}$		9.89	.006	$_{1,19}$	9.08	.007	-2.43	-3.14	.006*
${}_{Zf}W_{T2fl}+W_{Zf}$		9.67	.006	$_{1,19}$	8.99	.007	-2.42	-3.10	.006*
${}_{Zf}W_{T2ff}+W_{Zf}$		9.53	.006	$_{1,19}$	8.71	.008	-2.41	-3.08	.006*
${}_{Zg}W_{T2rr}+W_{Zg}$		5.32	.033	$_{1,18}$	5.22	.034	-1.93	-2.30	.033*
${}_{Zg}W_{T2xr}+W_{Zg}$		5.41	.032	$_{1,18}$	5.30	.035	-1.97	-2.32	.032*
${}_{Zg}W_{T2gr}+W_{Zg}$		5.41	.032	$_{1,18}$	5.29	.035	-1.95	-2.32	.032*

Interactions of condition by set type:

dep. variable		$F_{1,18}$	p	df	quasi-F	p	coeff.	t	p
${}_{Zf}W_{T2ll} - W_{Zf}$	❹$_d$	9.88	.006	$_{1,19}$	9.07	.008	2.42	3.14	.006*
${}_{Zf}W_{T2lf} - W_{Zf}$	❹$_d$	10.88	.005	$_{1,19}$	9.23	.007	2.43	3.16	.005*
${}_{Zf}W_{T2fl} - W_{Zf}$	❹$_d$	10.22	.005	$_{1,19}$	9.31	.007	2.44	3.19	.005*
${}_{Zf}W_{T2ff} - W_{Zf}$	❹$_d$	10.35	.005	$_{1,19}$	9.59	.006	2.45	3.21	.005*
${}_{Zr}W_{T2rr} - W_{Zr}$	❶$_e$	4.49	.048	$_{1,18}$	4.47	.048	1.87	2.11	.048*
${}_{Zg}W_{T2rr} - W_{Zg}$	❶$_e$	5.55	.030	$_{1,18}$	5.43	.031	1.97	2.35	.030*
${}_{Zg}W_{T2xr} - W_{Zg}$	❶$_e$	5.47	.031	$_{1,18}$	5.35	.034	1.93	2.33	.031*
${}_{Zg}W_{T2gr} - W_{Zg}$	❶$_e$	5.46	.031	$_{1,18}$	5.34	.034	1.96	2.33	.031*

Main effects of set type:

dep. variable		$F_{1,18}$	p	df	quasi-F	p	coeff.	t	p
$W_{Zl} - W_{Zf}$	❺$_d$	9.24	.007	$_{1,19}$	8.92	.007	4.19	3.03	.007*
$W_{T1} - W_{T2}$		55.08	.000	$_{1,17}$	36.71	.000	-.57	-7.42	.000*
${}_{Zl}W_{T2ll} - {}_{Zf}W_{T2ll}$	❻$_d$	21.94	.000	$_{1,14}$	16.09	.002	.28	4.68	.000*
${}_{Zl}W_{T2lf} - {}_{Zf}W_{T2lf}$		15.32	.001	$_{1,4}$	4.42	.100			
${}_{Zr}W_{T2rr} - {}_{Zg}W_{T2rr}$		12.49	.002	$_{1,18}$	12.91	.005	.06	3.53	.002*
${}_{Zr}W_{T2xr} - {}_{Zg}W_{T2xr}$		6.40	.021	$_{1,18}$	7.99	.015	.04	2.52	.021*
$W_{T1} - W_{Z}$	❶$_d$	306.25	.000	$_{2,18}$	303.45	.000	-19.97	-17.49	.000*
$W_{T2} - W_{Z}$	❶$_d$	303.06	.000	$_{1,18}$	297.87	.000	-19.40	-17.40	.000*
${}_{Zl}W_{T2ll} - W_{Zl}$	❼$_d$	160.58	.000	$_{1,19}$	147.13	.000	-12.47	-12.67	.000*
${}_{Zl}W_{T2lf} - W_{Zl}$	❼$_d$	157.02	.000	$_{1,19}$	143.92	.000	-12.70	-12.53	.000*
${}_{Zl}W_{T2fl} - W_{Zl}$	❼$_d$	156.58	.000	$_{1,19}$	143.77	.000	-12.71	-12.51	.000*
${}_{Zl}W_{T2ff} - W_{Zl}$	❼$_d$	156.21	.000	$_{1,19}$	144.12	.000	-12.73	-12.49	.000*
${}_{Zf}W_{T2ll} - W_{Zf}$	❼$_d$	123.11	.000	$_{1,19}$	111.52	.000	-8.57	-11.09	.000*
${}_{Zf}W_{T2lf} - W_{Zf}$	❼$_d$	123.81	.000	$_{1,19}$	112.28	.000	-8.56	-11.12	.000*
${}_{Zf}W_{T2fl} - W_{Zf}$	❼$_d$	124.89	.000	$_{1,19}$	112.19	.000	-8.54	-11.17	.000*
${}_{Zf}W_{T2ff} - W_{Zf}$	❼$_d$	123.62	.000	$_{1,19}$	111.92	.000	-8.49	-11.11	.000*
${}_{Zr}W_{T2rr} - W_{Zr}$	❷$_e$	54.41	.000	$_{1,18}$	44.66	.000	-6.53	-7.37	.000*
${}_{Zg}W_{T2rr} - W_{Zg}$	❷$_e$	53.97	.000	$_{1,19}$	47.80	.000	-6.15	-7.34	.000*
${}_{Zr}W_{T2xr} - W_{Zr}$	❷$_e$	55.44	.000	$_{1,18}$	45.38	.000	-6.56	-7.44	.000*
${}_{Zg}W_{T2xr} - W_{Zg}$	❷$_e$	55.26	.000	$_{1,19}$	48.58	.000	-6.16	-7.43	.000*
${}_{Zr}W_{T2gr} - W_{Zr}$	❷$_e$	55.21	.000	$_{1,18}$	45.69	.000	-6.60	-7.43	.000*
${}_{Zg}W_{T2gr} - W_{Zg}$	❷$_e$	53.82	.000	$_{1,19}$	47.64	.000	-6.15	-7.33	.000*

Interactions of condition by expression type:

dep. variable	Pillai's Trace	$F_{2,17}$	p
W_Z	.04	.41	.668

Main effects of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,17}$	p	df	quasi-F	p		coeff.	t	p
W_Z	.23	2.55	.107							
$W_{T1}+W_{T2}$ ❷$_d$	.50	8.53	.003	$_{2,23}$	4.04	.031	3	.30	4.12	.001**
${}_{Zl}W_{T2ll}+{}_{Zf}W_{T2ll}$	.39	5.46	.015	$_{2,38}$	3.24	.060				
${}_{Zl}W_{T2lf}+{}_{Zf}W_{T2lf}$ ❽$_d$	.43	6.47	.008	$_{2,36}$	8.03	.003	1	-.05	-3.68	.002**
${}_{Zl}W_{T2fl}+{}_{Zf}W_{T2fl}$ ❽$_d$	.41	5.96	.011	$_{2,26}$	5.32	.012	3	.06	3.48	.003**
${}_{Zl}W_{T2ll}+W_{Zl}$	.30	3.71	.046	$_{2,39}$	3.91	.026	2	2.14	2.80	.012**
${}_{Zl}W_{T2fl}+W_{Zl}$	.30	3.65	.048	$_{2,38}$	3.79	.041	2	2.05	2.78	.012**
${}_{Zl}W_{T2ff}+W_{Zl}$	.30	3.68	.047	$_{2,39}$	3.78	.032	2	2.05	2.79	.012**

Interactions of set type by expression type:

2 = set type vs (literals vs anomalies)
3 = set type vs (metaphors vs anomalies)

dep. variable		Pillai's Trace	$F_{2,17}$	p	df	quasi-F	p		coeff.	t	p
W_{T1} - W_{T2}	❸$_d$	.43	6.56	.008	$_{2,37}$	5.18	.032	3	-.25	-3.69	.002**
$_{Zl}W_{T2ll}$ - $_{Zf}W_{T2ll}$		.36	4.79	.022	$_{2,39}$	2.86	.072				
$_{Zl}W_{T2lf}$ - W_{Zl}	❾$_d$	.30	3.68	.047	$_{2,38}$	3.75	.037	2	-2.05	-2.78	.012**
$_{Zl}W_{T2fl}$ - W_{Zl}	❾$_d$	.29	3.59	.050	$_{2,39}$	3.65	.038	2	-2.03	-2.75	.013**

Table 5.8: MANOVA on grand mean similarity (*SIM*) and figurativeness (*FIG*) for judgements from '*A* to *B*' and '*B* to *A*' of literals (L), metaphors (M) and anomalies (A) in *expression* and *context* (N = 52).

	Pillai's Trace	$F_{1,48}$	p
Main effect of direction		2.16	.148
Main effect of condition		.06	.806
Direction by condition		.34	.561
Direction by dimension		.05	.831
Condition by dimension		.24	.629
Direction by condition by dimension		1.89	.175
	Pillai's Trace	$F_{2,47}$	**p**
Direction by condition by expression type	.02	.52	.594
Condition by expression type	.01	.41	.661
Direction by expression type	.08	2.18	.124
Direction by condition by dimension by expression type	.01	.40	.670
Condition by dimension by expression type	.02	.71	.492
Direction by dimension by expression type	.08	2.31	.110

Main effect of dimension:

dep. variable	$F_{1,48}$	p	df	quasi-F	p		coeff.	t	p
SIM - FIG	56.35	.000	$_{1,49}$	54.91	.000	1	-2.86	-7.50	.000*

Dimension by expression type:

1 = (*SIM* vs *FIG*) vs (literals vs metaphors)
2 = (*SIM* vs *FIG*) vs (literals vs anomalies)
3 = (*SIM* vs *FIG*) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{2,47}$	p	df	quasi-F	p		coeff.	t	p
SIM - FIG	.67	48.76	.000	$_{2,47}$	55.61	.000	1	2.86	9.66	.000**
							2	1.86	7.30	.000**
							3	-.99	-7.05	.000**

Main effect of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,47}$	p	df	quasi-F	p		coeff.	t	p
SIM+FIG	.81	102.25	.000	$_{2,10}$	49.44	.000	1	1.83	7.00	.000**
							2	4.35	13.44	.000**
							3	-.99	12.18	.000**

Table 5.10: MANOVA on the grand mean number of weighted features (words) per term presentation. Presentation 1 was performed in the preceding *single term*, whereas presentation 2 up to 4 were the repetitions in either *expression* or *context*. Main effects of condition, expression type, and interactions of expression type by condition are not printed. Figures in white circles followed by subscripts (e.g., ①$_f$) refer to effects discussed in the text.

Main effects of presentation:

		$F_{1,18}$	p	coeff.	t	p
$pres_{1-2}$	①$_f$	47.29	.000	7.20	6.87	.000*
$pres_{1-3}$	①$_f$	64.36	.000	6.74	8.02	.000*
$pres_{1-4}$	①$_f$	61.25	.000	7.23	7.82	.000*
$pres_{2-3}$		.61	.444			
$pres_{2-4}$		.00	.948			
$pres_{3-4}$		1.45	.245			

Interactions of condition by presentation:

		$F_{1,18}$	p	coeff.	t	p
cond*$pres_{1-2}$		.90	.356			
cond*$pres_{1-3}$		.04	.844			
cond*$pres_{1-4}$		.07	.793			
cond*$pres_{2-3}$		1.98	.177			
cond*$pres_{2-4}$	②$_f$	5.12	.036	-1.23	-2.26	.036
cond*$pres_{3-4}$		1.01	.329			

Interactions of presentation by expression type:

		Pillai's Trace	$F_{2,17}$	p	expr. type	coeff.	t	p
$pres_{1-2}$*type	③$_f$	.36	4.85	.021	L-M	-1.32	-1.60	.125
					L-A	-2.63	-3.15	.006**
					M-A	-1.31	-1.89	.074
$pres_{1-3}$*type		.26	3.13	.069				
$pres_{1-4}$*type	④$_f$	.33	4.25	.032	L-M	-1.63	-2.21	.040
					L-A	-2.70	-2.59	.018
					M-A	-1.06	- .98	.339
$pres_{2-3}$*type		.02	.24	.783				
$pres_{2-4}$*type		.01	.11	.888				
$pres_{3-4}$*type		.02	.18	.835				

Interactions of condition by presentation by expression type:

	Pillai's Trace	$F_{2,17}$	p
cond*$pres_{1-2}$*type	.07	.64	.536
cond*$pres_{1-3}$*type	.07	.72	.500
cond*$pres_{1-4}$*type	.10	.99	.391
cond*$pres_{2-3}$*type	.17	1.85	.187
cond*$pres_{2-4}$*type	.10	.97	.398
cond*$pres_{3-4}$*type	.01	.11	.895

Table 5.11: The significant effects of condition, expression type, term and word frequency on the grand means of (W_X , W_Y) and W_S (measure M2). White circled figures with sub-scripts index important effects discussed in the text.

Main effect of word frequency:

1 = high vs low

dep. variable	$F_{1,37}$	p		coeff.	t	p
$W_{SX}+W_{SY}$	5.19	.029	1	-.36	-2.28	.029*

Interaction of condition by word frequency:

2 = (*single term* vs *context*) vs (high vs low)

dep. variable	$F_{2,37}$	p		coeff.	t	p
$W_{SX}+W_{SY}$ ③$_g$	6.02	.005	2	-1.21	-3.33	.002**

Interaction of term by word frequency:

1 = (*A-term* vs *B-term*) vs (high vs low)

dep. variable	$F_{1,37}$	p		coeff.	t	p
W_X - W_Y ①$_g$	5.13	.029	1	1.03	2.27	.029*

Interactions of expression type by word frequency:

1 = (literals vs metaphors) vs (high vs low)
2 = (literals vs anomalies) vs (high vs low)
3 = (metaphors vs anomalies) vs (high vs low)

dep. variable	Pillai's Trace	$F_{2,36}$	p		coeff.	t	p
W_X+W_Y ②$_g$	.28	6.97	.003	1	-.74	-2.40	.022✔
				2	-1.07	-3.74	.001**
$W_{SX}+W_{SY}$ ④$_g$	.30	7.58	.002	2	-.33	-3.49	.001**
				3	-.27	-2.69	.011**

✔ = not significant at $\alpha \approx .0166$

Table 5.12: MANOVAs for the grand means of weighted feature set size (W_X , W_Y) and weighted shared set size W_S , evoked by either ambiguous (amb$_y$) or nonambiguous (amb$_n$) terms, in all expression types. White circled figures with subscripts (e.g., ①$_h$) direct to effects discussed in the text.

Interactions of condition by ambiguity:

2 = (*single term* vs *context*) vs (amb$_y$ vs amb$_n$)

dep. variable	$F_{2,37}$	p		coeff.	t	p
W_X+W_Y	3.76	.033	2	5.51	2.61	.013**

Interactions of ambiguity by expression type:

1 = (amb$_y$ vs amb$_n$) vs (literal vs metaphor)
3 = (amb$_y$ vs amb$_n$) vs (metaphor vs anomaly)

dep. variable		Pillai's Trace	$F_{2,36}$	p		coeff.	t	p
W_X+W_Y		.31	8.42	.001	1	-1.39	-4.06	.000**
					3	.72	2.56	.014**
W_S	④$_h$	.17	3.91	.029	1	.29	2.57	.014**
					3	- .27	-2.51	.016**

Interactions of condition by ambiguity by term by expression type:

2 = (*single term* vs *context*) vs (amb) vs (term) vs (literal vs metaphor)
3 = (*expression* vs *context*) vs (amb) vs (term) vs (literal vs metaphor)

dep. variable		Pillai's Trace	$F_{4,74}$	p		coeff.	t	p
$W_X - W_Y$	①$_h$	.31	3.41	.013	2	-2.04	-2.97	.005***
					3	-2.75	-3.47	.001***

Interactions of ambiguity by term by expression type:

1 = (amb$_y$ vs amb$_n$) vs (*A*- vs *B-term*) vs (literal vs metaphor)
2 = (amb$_y$ vs amb$_n$) vs (*A*- vs *B-term*) vs (literal vs anomaly)

dep. variable		Pillai's Trace	$F_{2,36}$	p		coeff.	t	p
$W_X - W_Y$	③$_h$	.33	9.08	.001	1	1.03	3.51	.001**
					2	.88	3.45	.001**

Table 5.14: MANOVA on the grand means of *A*- and *B-term* dominance for all expression types and conditions.

Main effect of condition:

1 = *single term* vs *expression*
2 = *single term* vs *context*
3 = *expression* vs *context*

dep. variable	$F_{2,37}$	p		coeff.	t	p
$A_{dom}+B_{dom}$	13.80	.000	1	-2.75	-5.18	.000**
			2	-1.36	-2.55	.000**
			3	1.40	2.27	.029✔

Interaction of condition by term:

1 = (*single term* vs *expression*) vs (*A*- vs *B-term*)
2 = (*single term* vs *context*) vs (*A*- vs *B-term*)
3 = (*expression* vs *context*) vs (*A*- vs *B-term*)

dep. variable	$F_{2,37}$	p		coeff.	t	p
A_{dom} - B_{dom}	8.22	.001	1	.37	2.48	.018✔
			2	- .33	-2.19	.034✔
			3	- .70	-4.05	.000**

Interaction of condition by expression type:

1 = (*single term* vs *expression*) vs (literals vs metaphors)
3 = (*expression* vs *context*) vs (literals vs metaphors)

dep. variable	Pillai's Trace	$F_{4,74}$	p		coeff.	t	p
$A_{dom}+B_{dom}$	.44	5.27	.001	1	1.28	4.30	.000***
				3	-1.01	-2.93	.006✘

Main effect of expression type:

1 = literals vs metaphors

dep. variable	Pillai's Trace	$F_{2,36}$	p		coeff.	t	p
$A_{dom}+B_{dom}$	.17	3.75	.033	1	- .34	-2.67	.011**

Interaction of condition by term by expression type:

dep. variable	Pillai's Trace	$F_{4,74}$	p
A_{dom} - B_{dom}	.15	1.55	.195

Interaction of term by expression type:

1 = (*A*- vs *B-term*) vs (literals vs metaphors)
3 = (*A*- vs *B-term*) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{2,36}$	p		coeff.	t	p
A_{dom} - B_{dom}	.17	3.94	.028	1	.09	2.40	.021✔
				3	- .15	-2.27	.029✔

✔ = not significant at α ≈ .0166
✘ = not significant at α ≈ .0055

CHAPTER 6: REACTION TIME*

6.0 The time course of metaphor processing

This is going to be a sad Chapter. Sad for those who are interested in metaphors, happy for those who are interested in methods. The metaphor models presented in the Chapters 1, 2, and 3 are not just a neat way to organize the activation and comparison of features, they also reflect assumptions on the order of processing. They are designed as serial phases (encoding, comparing, responding) and as serial stages. To understand a metaphor, the comparison model expects one stage and the anomaly model two stages (literal before figurative), just as the interaction model (normal features before relations). However, stages do not necessarily have serial orders. The results of the previous Chapter suggest that figurative features, for example, are produced in parallel with literal features (Figure 5.15, Chapter 5), and that response criteria may be checked synchronously. This Chapter is devoted to such issues as the order of processing the expression types, in one or two stages, serial or in parallel.

Whereas Chapter 5 researched the contents of the first two phases (encoding and comparing), Chapter 6 focuses on the response phase. According to the models, certain criteria for the size of the shared sets decide whether an expression is literal, metaphoric or anomalous. Literal expressions are supposed to be processed faster than metaphors, which are equally fast as anomalies.

Section 6.1 puts forward that reaction time (RT) experiments may be relevant for the study of literature, while Section 6.2 reviews such experiments on metaphor in psychology. This Section is also a minute and exhaustive review of methods and statistics, since Section 6.3 will argue that metaphor processing may be highly dependent of the experimental setting. Effects of metaphor on RT may be overpowered by nonlinguistic variables. In Section 6.4, an RT experiment is conducted that controls for various design artifacts, and that contrasts metaphors with simile variants. By interjecting the preposition 'like' ('als' in Dutch), subjects have the opportunity to make faster decisions for metaphors than for literals, which is not expected by any of the models. For metaphors, this should facilitate the process (faster and more correct decisions), whereas for literals and anomalies, it should be inhibited (slower decisions, tending to be 'metaphor'). Section 6.5 and 6.6 draw conclusions for psychology and the theory of literature on the purpose of RT in metaphor research.

---

* Notes are on page 258. Appendix with statistics on the pages 259-264.

6.1 What could reaction time say about literature?

Reading is time-consuming. Reading poetry is even more time-consuming than reading a paper article. Why? Poetry is more difficult. Poetry opens more options for polyvalence (Steen 1994: 142-143), due to greater lexical ambiguity, more metaphors, and a greater instance of novel words. A poetic context does not prime the meaning as much as a paper article does, so that readers search their memory more extensively, and exploit higher cognitive activity to create meaning.

> The technique of art is to make objects 'unfamiliar', to make forms difficult, to increase the difficulty and length of perception because the process of perception is an aesthetic end in itself and must be prolonged. (Šklovskij 1965: 12, originally 1917; Zwaan 1993: 41)

The prolonged process of perception is an aesthetic end in itself, in which the duration of processing a poetic phrase is longer. Since the understanding of what a poetic phrase means really takes time, real time *is* the relevant variable to study poetic meaning.

Two modes of language processing can be explored by measuring the time frame of sense creation: Automatization and foregrounding. According to Mukařovský (1964: 51-54), poetry needs the standard language - with its highly automatized phrasing - to make certain text elements conspicuous, to make them foregrounding. The automatized language forms the background to arrive at salient semantic deviations.

Understanding automatized language, then, should be a fast process. Most words are expected on the basis of the context (the selection from the paradigm is facilitated), so that the sequence becomes fluent and is easily understood.

By contrast, foregrounding elements are not primed by the surrounding text, and do not follow from the well-developed skilled behaviors of interpreting standard language. In a manner of speaking, readers 'stumble' over semantic deviations, are more aware of them, and invest more effort to integrate them in a meaningful frame. Of course, this reader-regulated processing mode is at the cost of spending more time on foregrounding elements.

The novel or inconsistent information that is characteristic for poetry thus should be represented by longer time intervals for meaning creation. Consider an example on foregrounding. A capitalized heading of a paper article is a backgrounding element for the genre, and its position at the top of the text is correct. Put in the middle of a poetic text, however, the same line becomes a foregrounding element, and its meaning will be contemplated longer. Thus, the dependency of foregrounding of genre specific contexts is reflected in the time to create meaning. Conversely, the time to give meaning to a texual element is sensitive to the interaction between genre and deviateness, between foregrounding versus backgrounding in paper articles versus poetry. Studying the understanding of poetry thus is the study of time.

Consider also the following. Anomalies hardly exist for sophisticated readers, *if they are given enough time* for sense creation. During this extra time, they can generate so much context that the anomaly becomes a metaphor. Thus, in single expression conditions, the anomaly takes more time to become a metaphor than in contexts that prime a metaphoric meaning of that anomaly. Again, time reflects the interaction between text condition and expression type, because anomalies become metaphors after contemplation, and contemplation takes time.

The problematic can be reduced to the following two positions. Poetry demands more control processing, because more concept-driven analysis is performed. Readers exhaust more long-term resources with respect to word, genre and world knowledge. Standard language involves more 'shallow' control processing. The data-driven analysis of text properties such as phonology, orthography and syntax is less disrupted by deviations, which do not need to be recovered by more knowledge based effort. From this perspective, the remark by Zwaan is most notable:

> One of the basic assumptions behind theories of literary reading is that contextual factors impinge upon meaning construction. It is perhaps no overstatement to say that most theorists of literary comprehension assign more weight to top-down processes in reading than to bottom-up processes. However, psycholinguistic studies of real-time processing show that the role of data-driven processes should not be undervalued. (Zwaan 1991: 169-170)

Thus, concept-driven analysis in poetry cannot be studied without taking account of data-driven analysis. It is precisely the deviant text properties in poetry which induce the higher activity of concept-driven analysis. In other words, a theorist of literary comprehension does not know what the concept-driven processes are, unless the effects of the text properties are captured. Measuring the real time of comprehension may be the way to do just that.

Two approaches for time measurement in literary comprehension are available: Reading times and reaction times. Reading time is the time to read a text from the first to the last word. They are expressed as milliseconds (ms) per word. Reaction time (RT) is the time from the onset of the stimulus (reading the *B-term*) to the initiation of the response ('it's a metaphor!'). Both variables may be informative for literary discourse analysis, as illustrated by the following example.

The first reaction time study by a literary theorist stems from Zwaan (1993). Reading times and reaction times were utilized to study the differences between processing literary and expository texts. Zwaan supposed that literature demands 'deep' processing, so that the reading times (ms/word) should slow down. Since they spent more time on the literary text, Zwaan supposed that readers paid more attention to the stylistics of the text, in other words, its surface form. Zwaan presumed that RTs for word recognition would be faster in the literary than in the nonliterary condition, since in the former, readers acquired more knowledge about the words themselves.

To test these deductions, Zwaan (1993: 51) presented two groups of subjects with 14 lines from a previously read text. The first group received a literary text, the second a journalistic text. Each line contained a capitalized word, which could be the original word from the previously read text or a synonym. Subjects pushed the response button 'correct' for the original words, whereas they pushed 'incorrect' for synonyms. The stylistic aspects were supposedly less indifferent to reading literature than to reading newspapers, so that the RTs for the literary group should be faster than for the journalistic group. However, no reliable differences in RT were found between the groups (Zwaan 1993: 56).

In other words, no evidence was found that the stylistic aspects were processed 'more deeply' in literature than in newspapers. It may be, then, that the theory was incorrect, or that the experiment should be devised differently. At least, by translating text differences into reaction time differences, a controllable variable was created for explicit assumptions on literary text processing, making the discussion more concrete. In psychology, reaction time experiments are common practice, also for the inquiry into (literary) metaphor processing.

6.2 RT studies on metaphor in psychology

This Section reviews a sample of RT experiments on metaphors, which are key studies in the area. They are usually referred to as 'valid claims' on metaphor processing. Since the quality of an argument depends on the quality of the experiment, the emphasis will be on designs and statistics. Potential critique will be presented in Section 6.3, so as to design a new RT experiment in Section 6.4. A more inclusive and less technical review was given by Hoffman & Kemper (1987). Gibbs (1994) provided a comprehensive review of the area. Note that most RT-studies are concerned with tests of the serial two-stage anomaly model.

Ortony, Schallert, Reynolds & Antos (1978) carried out two RT experiments, in which contexts primed the literal or metaphoric meaning of metaphoric and idiomatic expressions. Idioms (e.g., 'a pain in the neck') are metaphors that have become so conventional that their literal meaning is subordinate. In experiment 1, the time was recorded to understand an expression. Context length (short vs long) was compared with prime type (cuing the literal or metaphoric meaning of an expression). According to a serial two-stage anomaly model, expressions primed as metaphors were supposed to be processed more slowly than those primed as literals. Moreover, it was expected that this difference would only occur in short contexts. Longer contexts were supposed to provide so many cues, that the relevant stage was evoked directly and, hence, any difference would be eliminated. Thus, the seriality of the two-stage model was supposed to be confirmed in short contexts, not in long ones.

The design combined two lists of expressions, two presentation orders and two contexts (short vs long) as between-subject factors. Literal or metaphoric priming was a within-subject factor.

Presentation order had no effect, and was excluded from further analysis. Unfortunately, a potential effect of expression list was not reported. Analysis of variance of the remaining factors indicated that metaphoric interpretations took longer than literal interpretations in the short contexts, whereas they took about equal time in the long contexts. As will be clear at the end of this Section, this is virtually the only RT experiment on metaphor processing that can bear the scrutiny of criticism.

Ortony et al. (1978) conducted a second experiment in which the manipulation was more evasive (cf. Hoorn, in press). Again, the time taken to understand an expression was recorded. Three between-subject factors of expression type order (idioms first or last), expression list and presentation order were employed in combination with two within-subject factors: Context type (long vs short) and expression type (idioms vs literal). The between-subject factors proved insignificant, and need not concern us here.

The number of expressions in the two within-subject factors of context type and expression type were not counterbalanced, so that their effects were confounded with practice. Two stories served as contexts (story A and B), while expressions could be idiomatic or literal. However, the idioms were connected to both stories, whereas the literal expressions were connected to story A only. Thus, the missing cell was literal expressions in story B. The metaphoric or literal priming was also unbalanced. Story A was constructed such that it primed the metaphoric meaning of the idioms, and the literal meaning of the literal expressions. Story B, however, only primed the literal meaning of the idioms. In other words, apart from prime type, another factor of story type should have been devised, so that story A and B would have primed the literal or metaphoric meaning of idioms and literal expressions. Another solution would have been to omit one of the stories, or to weigh the results of story A as half.

These studies by Ortony, Schallert, Reynolds & Antos (1978) initiated a series of RT experiments on metaphor and idiom processing. For instance, Gibbs (1980) conducted an experiment on idioms ('he's singing a different tune'). As mentioned above, idioms are ambiguous stimuli, because they have a literal and a metaphoric meaning. The metaphoric meaning has become conventional in the standard language, whereas the literal meaning is hardly ever used. Gibbs investigated whether the conventional metaphoric meaning was processed faster than the unconventional literal meaning, even when the literal meaning was primed by the context.

In one subject group, idioms were used as prime, which were followed by literal or metaphoric paraphrases of these idioms. In the second subject group, idioms were embedded in contexts, cueing either the literal or metaphoric meaning of the idiom. Here also, literal or metaphoric paraphrases followed the idioms.

To make sure that the first group would not interpret the idioms only in their metaphorical sense, the idioms were mixed with literal expressions, which also received literal or metaphoric paraphrases. RT was registered for true-false decisions after paraphrase presentation. The results suggested that conventional (metaphoric) uses of idiom were processed faster than unconventional (literal) uses, despite their metaphoric origin and despite priming the literal meaning.

To summarize, idioms and literal expressions without context preceded paraphrases that could be either literal or metaphoric, and idioms in a literal or metaphoric context preceded paraphrases that could be literal or metaphoric. Gibbs analyzed this design as a standard 2*2 of condition (context vs no-context) and paraphrase (literal vs metaphoric). Yet, the design was a 2*2*2 of condition, paraphrase (literal vs metaphoric), and prime type (literal prime vs metaphoric prime).

To know whether the prime type of the contexts was literal or metaphoric, a new group of subjects rated the priming of the contexts, whereas the priming of the idioms and literal expressions was not rated. Thus, the prime type in no-context might not have been the same as the prime type in context. Yet, they were treated as comparable in the design, whereafter only the effects of idioms - not of the literal expressions - were analyzed.

Important information was lost concerning idioms with literal primes and metaphoric paraphrases, and those with metaphoric primes and literal paraphrases. These could have told the power - within subjects - of the literal and metaphoric primes in the interaction of (condition by) prime type by paraphrase.

Glucksberg, Gildea & Bookin (1982) administered a within-subject experiment on metaphors and literals in a standard sentence verification task (true/false). Literals could be correct ('Standard True') or incorrect ('Standard False'); accordingly, metaphors could be correct ('Metaphors') or incorrect ('Scrambled Metaphors'). The latter can be viewed as instantiations of anomalies. Subjects judged the truth value of the expressions, and the RT from expression onset to button press was measured.

It was argued that if the figurative meaning of correct metaphors can be ignored, 'false' decisions should be equally fast for correct and incorrect metaphors. Conversely, if the figurative meaning is accessed automatically, 'false' decisions for correct metaphors should be slower than for incorrect metaphors, because an extra figurative check is needed. It was found that RT for correct literals was fastest - followed by incorrect literals and incorrect metaphors - whereas correct metaphors were slowest. It seemed that metaphors were evaluated on their figurative truth, which is consistent with the two-stage model. Incorrect literals and incorrect metaphors were perceived as equally anomalous and could be rejected as not literal.

However, the study has a number of pitfalls. The number of literals and metaphors were not balanced. The correct literal expressions had 80 items, the incorrect 40, the incorrect metaphors had 20 items, and the correct metaphors also 20. As indicated above, incorrect literals and incorrect metaphors took about an equal time to be rejected as 'false'. In other words, the incorrect expressions may all have been perceived as anomalies. If so, it could be that subjects perceived 80 literals, 60 anomalies and 20 metaphors. This range coincides with the ordinal pattern that was found in the mean RTs. Thus, reactions may become slower as expressions occurred less frequently.

The authors suggested that the semantic relationship between *A-* and *B-term* was probably less strong for metaphors than for literals, so that metaphors were processed more slowly. If so, the authors argued, the RT pattern should remain unchanged even in different contexts.

In experiment 2, expressions were provided with context by introducing the quantifiers *Some* and *All*. '*Some* surgeons are butchers' was supposed to be more plausible than '*All* surgeons are butchers'. The more plausible or 'correct' metaphors were rated for 'goodness', whereas the other expression types were not. In the interaction of expression type (literal vs metaphor) by quantifier, the *Some* metaphors (rated as 'good') showed the pattern of experiment 1 that RT for correct literals was fastest - followed by incorrect literals and incorrect metaphors - whereas correct metaphors were slowest. The *All* metaphors (rated as 'less good') did not show this pattern. Thus, context had effect, and it was inferred that metaphor comprehension was not merely a matter of semantic relatedness.

This experiment was analyzed as two quantifiers with four expression types. The expression types (Standard True, Standard False, Metaphors, Scrambled Metaphors) were treated as four independent conditions. Yet, an important interaction with correctness was overlooked. Standard True and False are literally correct and incorrect expressions. Metaphors and Scrambled Metaphors are figuratively correct and incorrect. Thus, the design is a 2*2*2 MANOVA for quantifier ($\exists$ vs $\forall$) by expression type (literal vs metaphor) by correctness (correct vs incorrect). Differences attributed to expression type may be due to correctness in the interaction. The same is valid for experiment 3, which duplicated this design and analysis.

Gildea & Glucksberg (1983) administered three variations on the above study, which combined three prime types (literal prime, figurative prime, no-prime) with four expression types (Standard True, Standard False, Metaphors, Scrambled Metaphors). Despite this 3*4 design, Standard True and False were not analyzed in experiments Ia, Ib and II, while the remaining 3*2 was analyzed only by planned comparisons (t-tests). Ib was analyzed in a 2*2 ANOVA, ignoring the higher-order interactions outlined earlier.

Estill & Kemper (1982) surveyed idioms (e.g., 'climbing the walls') for effects on RT with three cue types and four context types. Subjects were asked to identify the last word ('walls') in an expression, given a particular precue. Cues could be words ('walls') identical to the last word in the expressions, they could be rhyming ('falls') with the last word or they formed the semantic category ('part of a building') of the last word. The idioms appeared in literal, figurative, or ambiguous contexts. Expressions that were not idioms but did use the idiom's last word ('knocking out the walls') were presented in nonidiomatic contexts. Subjects reacted as soon as the last word of an expression was encountered in the context.

This 3*4 design was properly analyzed and it was found that the main effects of cue type and context type were significant, whereas the interactions were not. The Word Identity cue yielded faster responses than the Rhyme cue, which was faster than the Semantic Category cue. More importantly, nonidiomatic contexts slowed down RT, compared with all other contexts. Thus, the last word was differently processed when it was part of an idiom than when it was not. The authors suggested that idioms were processed as discrete lexical entries.

There is a slight inconvenience, however, concerning the last words of the expressions. In the nonidiomatic context, the last word came from an idiom - but was not used in an idiom - as opposed to the other contexts. Thus, the stimulus ratios were not counterbalanced, since three idioms were contrasted with one nonidiom. This may explain the slower RTs for nonidiomatic contexts. Since the argument of automatic access of idiomatic meaning was entirely based on this RT difference, the effects should have been treated more carefully.

Unfortunately, the authors did not provide the exact p-values for the t-tests on context types. However, the definition of the α-level is critical for the acceptance of - in this case - automatic access of idiomatic meaning. Context type had four levels (literal, figurative, ambiguous, non-idiomatic), and thus established six related main effects. In other words, the p-value should not have been tested with $\alpha = .05$, but rather with $\alpha = 8.33^{-03}$ (cf. Hoorn in press). Likewise, if the interactions had been significant, three cue types by four context types would have resulted in 18 related interactions, and should have been tested with $\alpha = .05 / 18 = 2.77^{-03}$.

Gerrig & Healy (1983) examined within-subject effects of prior and ensuing contexts on reading time for metaphors. Metaphors rated as 'good' were contrasted with 'bad' metaphors in active and passive voice sentences, which coincided with prior and ensuing context: 'The train followed the parallel ribbons' vs 'The parallel ribbons were followed by the train'. Subjects pushed a button as soon as they understood the sentence. Only the main effect of context was significant, indicating that prior context yielded faster reading times than ensuing contexts, or - as a confounded alternative - confirming the classic finding that active voices were processed faster than passive voices.

Each cell in the analysis contained 8 metaphors. However, the experiment also employed 16 fillers. In other words, the stimuli were not counter-balanced (32 metaphors against 16 fillers). Fillers were not rated on potential metaphoricalness, so that their effects were unpredictable. Filler effects were not analyzed together with the metaphor effects in a (weighted) MANOVA. Since only one random order was used for the stimuli, all subjects received stimuli in fixed order, which may easily affect the data. The same applies to the second experiment.

As indicated by the authors, the manipulation confounded ensuing context with passive voice. Therefore, another within-subject design contrasted literals with metaphors in active and passive voice. It was argued that if slow reading was induced by passive voices rather than ensuing contexts, this effect should equally occur for metaphors and literals. If, on the other hand, metaphors increased reading times in passive voice - whereas literals would remain equal - the effects in experiment 1 could be attributed to ensuing context, not to passive voices. Significant interactions between (literal-metaphor) and (active-passive voice) that increased metaphor reading times would underscore the idea that ensuing context played a special role in metaphor comprehension.

Indeed, such a significant interaction was found. However, the authors indicated that the literal expressions had shorter sentences than metaphors. They asserted that this might have influenced the relative speed with which literals and metaphors were read, but that this was immaterial for their argument. Thus, the main effect of expression type was irrelevant, and only the interaction counted. Nonetheless, the significant interaction may not only indicate that reading times for metaphors were increased by the ensuing contexts, but also that short sentences (literals) canceled the elongating effect of passive voices. If so, the interaction of expression type and context is confounded again.

Inhoff, Lima & Carroll (1984) verified the results of Ortony et al. (1978) with three experiments. In experiment 1, short contexts primed either the literal or metaphoric meaning of metaphors, whereas in experiment 2, long contexts did. Both experiments also employed unrelated contexts, which were inappropriate combinations of context and metaphor. Sentence reading time and total viewing time on critical words were measured with an eye tracker. The authors claimed that metaphorical meanings were understood as quickly as literal ones in long contexts, but slower in short contexts.

Stimulus lists were varied between subjects, while related vs unrelated context was varied within subjects. Context could prime the literal or metaphoric meaning. Moreover, short and long contexts were compared across experiments. Thus, the complete design was a 2*3*2*2 MANOVA for experiment by list by relatedness by prime type. Yet, two separate analyses for experiment 1 and 2 merely tested three means for the main effects of prime type: Literal vs metaphoric vs unrelated. The discussion suggested that there was an interaction of prime type with context, but without any supporting statistical tests.

Additionally, the metaphors were presented in the same serial order for each list, so that practice or boredom effects were not recognized. Furthermore, the priming effect of literal and metaphoric contexts was scored for the appropriate metaphors. However, it was not for the inappropriate ones.

Experiment 3 investigated the thematic relatedness between context and metaphor on reading time. The short contexts of the previous experiments served as the thematically related condition, whereas newly created 'associated-words' contexts served as the unrelated condition. Prime type was literal or metaphoric in both conditions, while expression type could be literal or metaphor. Six lists were a between-subject factor, whereas relatedness, prime type and expression type were varied within subjects. Thus, the design was a 6*2*2*2 MANOVA for list by relatedness by prime type by expression type.

Nonetheless, the design was diagnosed as 6 lists by 3 prime types (literal vs metaphoric vs associated-words) by 2 expression types, and was yet analyzed by 2*2 ANOVAs for e.g., prime type (literal vs metaphoric) versus expression type (literals vs metaphors).

Paivio & Clark (1986) explored the importance of imagery and intelligibility for the processing time of metaphors. Three prime types were used between subjects: A cue for the *A-term*, a cue for the *B-term*, and no cue. Metaphors were scored within subjects for the imagery value of the *A-term*, of the *B-term*, and for the intelligibility of the metaphor. Subsequently, these metaphors were orthogonally distributed over high or low *A-term* imagery, *B-term* imagery, and intelligibility. Subjects read the metaphors after the presentation of a prime, and released a button when they were ready to paraphrase the metaphor.

In principle, this is a 3*2*2*2 MANOVA for prime type by *A-term* imagery by *B-term* imagery by intelligibility. Yet, prime type was analyzed as a main effect of three individual means, separately from the three scale factors, which were analyzed as a second-order interaction. The main effect of prime type was only significant with stimuli as random factor, so that it was merely an idiosyncrasy of the subject group.

The authors noticed that the results were paradoxical. Since the *B-term* is the image in which the *A-term* is perceived, *B-term* imagery should be and actually was a stronger correlate of metaphor comprehension time than *A-term* imagery. However, the expected superiority of priming the *B-term* over priming the *A-term* or no priming did not appear. This remark seems curious, given the fact that the third-order interaction was never analyzed.

Janus & Bever (1985) also verified the results of Ortony et al. (1978). Literal contexts primed literal meanings, whereas metaphoric contexts primed metaphoric meanings of idioms. The time was measured that subjects indicated that the idiom was understood. RT was slower for metaphoric than for literal meaning, which was interpreted as support for the two-stage anomaly model.

Nonetheless, idioms are ambiguous expression types. Actually, they consist of two expression types, i.e. a literal expression and a metaphor, which are easily activated by the appropriate context. The authors indicated that the idioms in the metaphoric contexts were consistently judged as less predictable than the idioms in the literal contexts. They attributed this effect to the literal meaning of the idioms. However, it was overlooked that the metaphoric meaning was systematically correlated with one set of texts, whereas the literal meaning was systematically correlated with another set of texts. In other words, since they used two different text sets to prime one of the meanings, they could not determine whether the effects were due to divergent idiom meaning or to different success in priming.

Glucksberg, Brown & McGlone (1993) studied whether conceptual analogies motivated the use and comprehension of idioms in discourse. Two reading time experiments were performed, in which two contexts were combined with two idiom types. Contexts indicated which person (S or C) in the story was the referent of the idiom. The idioms differed for their consistency with the stories (consistent vs inconsistent). Subjects read the stories for comprehension and indicated that they finished reading by pressing a key.

However, the actual manipulation looked different. Their Appendix B lists two different contexts, two different idioms ('blew top' vs 'bit head off'), two different actors ('S blew top' - 'S bit head off' vs 'C blew top' - 'C bit head off') and two different referents ('S blew S's top' - 'S bit C's head off' vs 'C blew C's top' - 'C bit S's head off'). Obviously, this is not the 2*2 ANOVA that the authors suggested, but rather a 2*2*2*2 MANOVA of context, idiom, actor and referent. Moreover, context 1 was systematically correlated with the combination SS and SC, whereas context 2 was systematically correlated with CC and CS. Thus, the reading time effects for context by idiom were biased by the systematically correlated interaction with actors and referents.

As a general remark, the studies reviewed above mainly explored the two-stage anomaly model with idioms. However, idioms are both literal and metaphoric expressions that share the surface form, which makes them improper stimuli to test the two-stage model. First, using idioms presupposes that the two-stage model is context-dependent, otherwise the manipulation with literal and metaphoric primes would be ineffective. Second, *any outcome based on idioms is congruous with a context-dependent two-stage model*. If idioms are fast with literal primes, they were supposedly perceived as a literal expression, so that a second figurative stage was not necessary. If idioms are slow in literal contexts, they were perceived as metaphors, so that figurative interference took place. If fast in metaphoric contexts, they were metaphors, the figurative stage of which was entered immediately. If they are slow in metaphoric contexts, it mainly shows that context was unimportant to the two-stage model, since the literal stage was completed before the second was initiated.

Thus, before including idioms in the stimulus set, it should be demonstrated that the two-stage model is sensitive to context. This could be done by crossing purely literal and purely metaphoric expressions with literal and metaphoric contexts.[1] If the two-stage model is context-independent, one effect should be significant: A main effect of expression type should underscore that literals are always faster than metaphors. If the model is context dependent, the literal context shows that literals are faster than metaphors, which effect should change in metaphoric contexts. The direction of the change in the metaphoric context is the interesting part. If metaphors are less slow in metaphoric than in literal contexts, then indeed there are two stages, the second of which is facilitated by the proper prime. If metaphors are as fast as literals in metaphoric contexts, then the two serial stages in literal contexts become parallel in metaphoric ones. If metaphors are faster than literals in metaphoric contexts, there are not two stages within one process, since either a literal stage or a figurative stage was executed, dependent on the context prime. All other patterns are also evidence against the two-stage model.

Idioms can be used only in combination with purely literal and purely metaphoric expressions. They are not suitable to test a metaphor model, because they are ambiguous. Idioms may only be used to investigate the conventionality of their literal or metaphoric meaning. Dependent of the pattern they follow in literal and metaphoric contexts, they may be processed as literal or metaphoric expressions and pass through one or more stages. If they behave as literal expressions, idioms have become conventional uses of figurative meaning. If they behave like metaphors (perhaps in metaphoric contexts), the conventional use is affected by figurative interference.

Similes (comparisons with the preposition *like*) are a more suitable expression type. In contrast with metaphors (*A* is *B*), the preposition (*A* is like *B*) may immediately launch the figurative or relation stage, so that the fixed seriality of stages is limited by the linguistic indicator *like*. In that case, decisions for literal expressions in simile version ('the sun is *like* a star') should be more confusing and should take longer than normal. Decision errors for literals and anomalies with the preposition should be directed to 'metaphor' more than without. These ideas will be tested in Section 6.4.

In this Section, leading reaction-time studies on metaphor processing were evaluated. It was found that most of the results are unreliable, owing to design artifacts and careless analysis. It was also argued that similes rather than idioms should be exploited to examine the metaphor models. Ultimately, claims such as those made by the review of Hoffman & Kemper (1987), that much is known about the time aspects of metaphor processing are over-optimistic. Their observation that much has yet to be learned is more appropriate. Since what is known about metaphor processing is based on weak experiments and inaccurate analysis, the conclusions of these studies remain highly speculative.

6.3 What should an RT study on metaphor provide?

Theoretical discrimination

The reaction time research reviewed above concentrated on the two-stage anomaly model. However, the comparison and interaction models have disparate explanations for the order of processing, so that RT results should not be interpreted solely from the standpoint of the anomaly model. Experiments should provide a possibility to compare the rival models.

A big problem in this matter is that the models do not allow for different RT patterns. They all predict that literal expressions are fastest, followed by metaphors and anomalies. If the inference rule is applied that criterion checks on the size of the shared set are serial (fixed order), and faster than the summated durations of activation and comparison, the comparison model does differ from the anomaly and interaction model. The time span between literal expressions and metaphors should be much longer in the latter two models, because the corresponding decisions are separated by a new stage, not just by a criterion check. However, how much longer this difference should be, is an empirical question and cannot be derived from the models.

It is undesirable to have identical predictions for competing models. Fortunately, the anomaly model offers possibilities for differentiation. When expressions do not satisfy the criterion for the literal shared set, a covert effect of anomality occurs - the shock effect - which invokes the second stage. In Figure 2.1, Chapter 2, this was represented by the box 'conclude anomaly'. This covert effect may be replaced by another overt criterion response. If an expression is not literal, it may be checked whether the shared literal set is so small that further processing is abandoned and the decision becomes 'anomaly'. On the other hand, if this lower boundary is surpassed, the anomalous moment remains covert and the second stage is entered. In this case, two different mean reaction times for anomalies can occur: A fast one, situated between decisions for literals and metaphors, and a slow one, equal to metaphors.[2] Thus, the anomalies may stop the process after they have been dismissed as literals, or they stop the process, when reconsidered as metaphors and yet rejected.

Controls on design artifacts

To enable theoretical conclusions, literals and metaphors ought to be contrasted with anomalies. Anomalies are decisive, because their position in the ordinal RT pattern sets apart the anomaly model (literal < anomaly < metaphor ≈ anomaly) from the comparison and interaction model (literal < metaphor ≈ anomaly). The latter two models cannot be distinguished, because the duration difference between literal expressions and metaphors remains unknown.

The introduction of anomalies induces a major problem that has never been recognized in metaphor research, namely the choice of choice task. If the three expression types are tested in combination, a 3-choice task (literal-metaphor-anomaly) causes difficulties that are not found in a 2-choice task, in that subjects may decide to ignore one option and concentrate on the other two. A 2-choice task for three decisions requires three experimental runs: L-M, L-A, M-A. In that case, metaphors may be faster or slower simply as an effect of the combination with literals *or* anomalies, whereas they would not if they were accompanied by literals *and* anomalies, as in the 3-choice task. Moreover, three 2-choice tasks repeat stimuli, whereas the 3-choice task does not. If repetition affects processing time, the 2-choice tasks may fail to show evidence for a certain model, whereas the 3-choice task does.

Either kind of choice task is arbitrary. The results of an experiment may be fully limited to the specific design, which could count as a reproach leveled at the RT studies on metaphor thus far. Choice task should be treated as a factor in the analysis. Metaphor research has devoted a great deal of attention to context effects, which requires a baseline condition of no-context (i.e. *expression*). However, little attention has been paid to the effects of condition order (presenting *context* before *expression* or vice versa). The same applies to the order of choice task (3-choice before 2-choice and vice versa). When repeated presentation infringes on metaphor processing, the models can be discharged from their claims on the time relations, or it may simply be that RTs are not the best means to investigate them.

Thus, precise evaluation of the designs and statistics in metaphor research is not just a drill in methodological hair-splitting. The metaphor models structure the mental chronometry of figurative information processing, which is a highly sensitive operation. It could easily be wiped out by effects of repetition, order of presentation or choice task.

6.4 **RT experiments for metaphors and similes**

Twenty-four undergraduates of language and literature, Dutch native speakers, aged between 18 and 28 years took part in a classification decision task (cf. Hoffman & Kemper 1987: Table 1). They were paid volunteers who had never previously participated in an experiment, and were uninformed about stimuli or research aims. Subjects were seated 50 centimeters in front of a computer screen and all had normal or corrected to normal eye sight. The experiment was programmed with the software package ERTS (Beringer 1992), running on a 486DX2.

Stimuli were the expressions of Table 4.1 (Chapter 4) from which an extra set of simile variants were constructed. Thus, 'the sun is a grape' had a simile counterpart 'the sun is like a grape'. The same was done for literal and anomalous expressions.

Trials consisted of a sequence of prime - fixation line - target. In the *expression* condition, the prime consisted of an expression without *B-term* (the sun is a), after which a fixation line was projected (the sun is a ____). The width of the fixation line had four or five underscores, corresponding to the number of *B-term* letters. Targets were *B-terms*, completing the expression (the sun is a grape). Trials in the *context* condition were composed of similar sequences, except that the prime was extended with the original poem.

Reading the primes was self-paced and finished by pressing the space bar of a standard key board. The fixation line was presented for one second, whereupon it was substituted by a two-second presentation of the *B-term*. *B-terms* could be either literal, metaphoric or anomalous. Upon *B-term* presentation, subjects made a speeded decision on the status of the expression: Literal, metaphor or anomaly.

Two choice tasks were explored: A two and a three choice decision task. For either choice task, the space bar functioned as the home key, carrying a green mark for the index finger to be returned to after response execution. The decision keys were the d, the 7 and the l, on equal distances from the green mark. All other keys were removed from the keyboard. Decision keys were marked with either an L, M or A, conform the choices 'literal', 'metaphor' and 'anomaly'. In the three-choice task, subjects received all expression types and decided among the three keys. The two-choice task implemented three combinations of two keys (e.g., L-M, L-A and M-A) by removing the superfluous key and mixing the two suitable expression types per key combination.

Subjects were instructed to handle home key and decision key only with the index finger of their preferred hand. They were instructed to release the home key when their decision was completed. Anticipatory releases automatically aborted the trial, and subjects received a warning. The trial was repeated later in the session. Decision times longer than two seconds also triggered a warning, and were discarded from analysis.

Prior to the experiments, subjects practiced in the three-choice task with stimuli unrelated to the test set, until the standard deviation of the reaction times was less than 10% to 15% of the mean (Sanders 1980). The practice trials were repetitions of 9 prototypical exemplars from each expression type (e.g., 'love is a rose'), according to a norm group. Errors were fed back by a warning and the correct answer. Decisions on practice trials that exceeded two seconds after *B-term* presentation also received warning feedback. Consequently, subjects were trained to react fast in the hardest choice procedure, and were stipulated by the prototypes for expression type definition. Decision corrections were not provided in the actual experiments.

The design had four between-subject factors, and three within-subject factors. The between-subject factors were preposition ('without like' vs 'with like'), task order (3-choice before 2-choice vs 2-choice before 3-choice), condition order (*expression* before *context* vs *context* before *expression*) and decision key distribution (LMA vs ALM vs MAL vs MLA vs AML vs LAM). The presentation orders of decision key combinations in the 2-choice task were for LMA: LM-LA-MA and MA-LM-LA; for ALM: AL-AM-LM and LM-AL-AM; for MAL: MA-ML-AL and AL-MA-ML; for MLA: ML-MA-LA and LA-ML-MA; for AML: AM-AL-ML and ML-AM-AL; for LAM: LA-LM-AM and AM-LA-LM. Note that the decision key distribution in the 3-choice task and the presentation order in the 2-choice task were not analyzed as separate factors, because zero degrees of freedom would remain for the residual error term. Since these were counterbalanced variables, they were pooled over subjects.

The three within-subject factors were choice task (3-choice vs 2-choice), condition (*expression* vs *context*) and expression type (literals vs metaphors vs anomalies), thus yielding 18 values per subject. Choice task and condition were blocked, whereas expression types were mixed within blocks. All factors were counterbalanced over subjects.

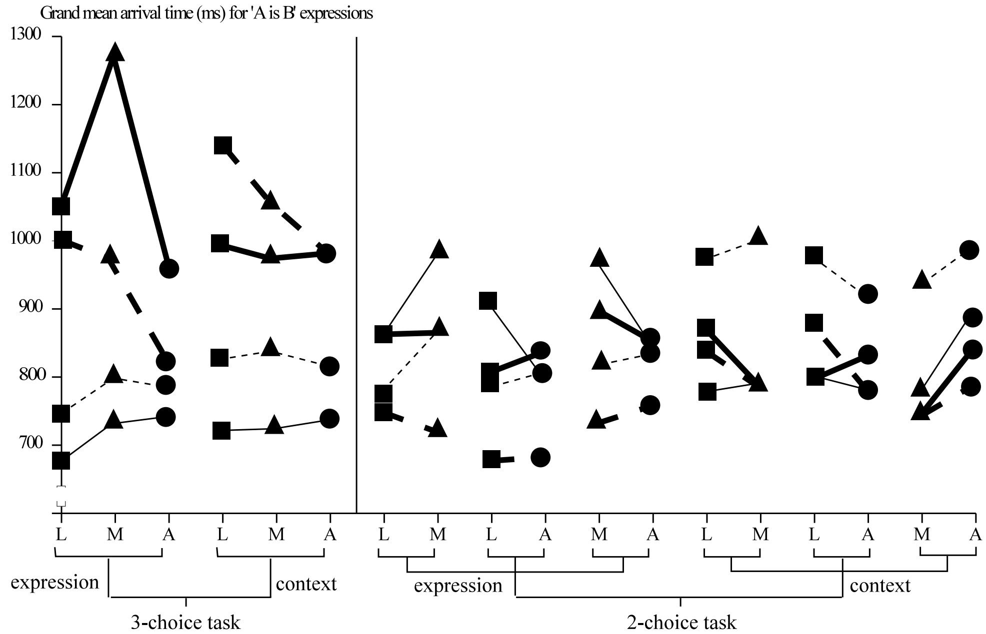
For the 3-choice task, 27 trials were distributed over 9 expressions per expression type, whereas the 2-choice task distributed 18 trials over two expression types of 9 expressions each. Each subject received different variants of the expression types. Variants of the same root were never clustered, so that in the overall design, each expression was used four times in different constellations. Within subjects, the stimuli in 2- and 3-choice, *expression* and *context* were the same, whereas they differed between subjects. No repetitions of *B-terms* occurred in a block of trials. Stimuli were presented pseudo-randomly, never offering more than two stimuli from the same expression type in a row. Upon *B-term* onset, RT was measured to the nearest millisecond, the dependent variables being release time from the home key (RT), movement time from home to decision key (MT) and decision accuracy.

MANOVA was run on the grand mean RT and RT+MT for correct decisions in a 2*2*2*2*2*3 design of preposition, task order, condition order, choice task, condition and expression type. Since the 3-choice task used two times 27 trials, whereas the 2-choice task used two times 54 trials, the transformation matrix weighted the values for the 2-choice task but half. Univariate tests were performed to calculate quasi-F values, for which - in view of the different weighing - the transformation matrix was submitted to Gram-Schmidt orthonormalization.

The problem arose that the effects of preposition, expression type, and condition on RT differed from RT+MT, which should not have been the case when subjects indeed released the home key after completion of the decision. MT is supposed to be a time constant needed only for motor execution. Thus, despite urgent instructions, the movement time between home key release and pressing the decision key was used to complete the decision. Therefore, only the results of RT+MT shall be reported, referred to as RT_a (arrival time). Nevertheless, the objection remains that MT may reflect cognitive processing in certain conditions, whereas in others, it might not.

Results for arrival times

Figure 6.0: Grand mean arrival times in milliseconds (ms) for correct responses to literal expressions (■), metaphors (▲) and anomalies (●) with or without the preposition 'like'. Bold lines indicate 3-choice before 2-choice task, thin lines 2-choice before 3-choice. Solid lines symbolize *expression* before *context*, dashed lines *context* before *expression*. In each cell, N = 3.



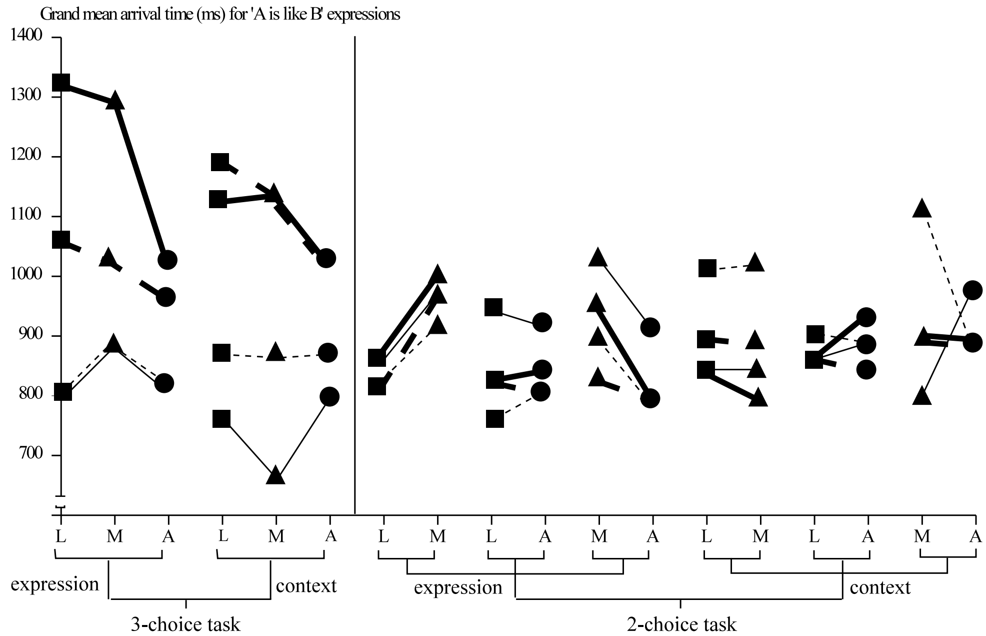
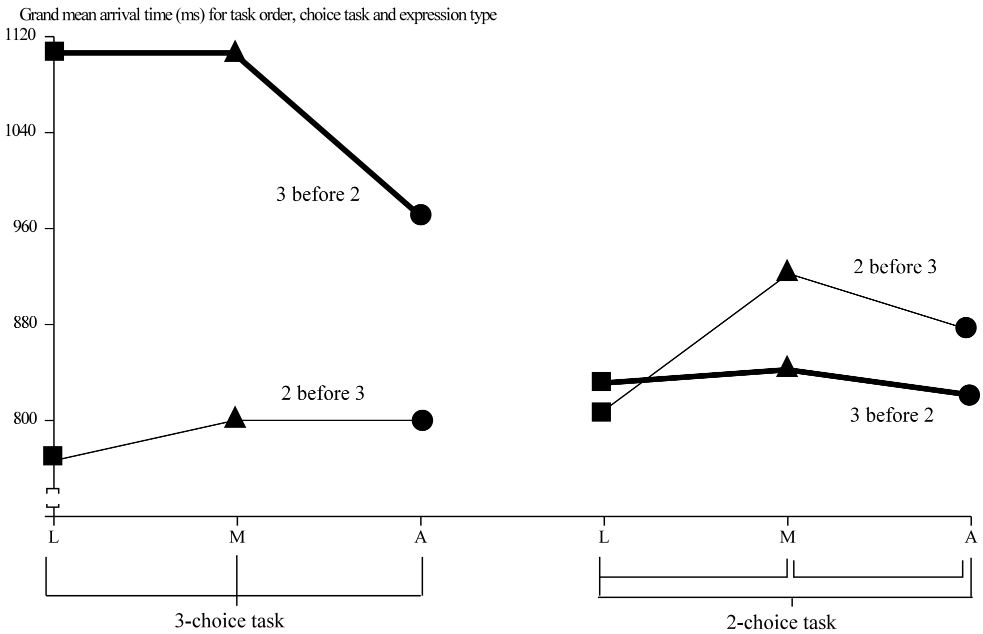


Figure 6.0 shows the grand mean arrival times for the correct responses to '*A* is *B*' (upper panel) and '*A* is like *B*' constructions (lower panel). It is evident that the results are quite contradictory and sensitive to various kinds of manipulation.

Table 6.0 (Appendix) shows that the only effects significant for both subjects and stimuli as the random factor are the main effect of expression type, the interaction of task order by expression type, and the interaction of task order by choice task by expression type. As a summary of Figure 6.0, Figure 6.1 depicts this higher-order interaction, from which the following might be construed:

(I) The overall response time was longer when 3-choice was presented before 2-choice than when 2-choice was presented before 3-choice. (II) Both 3- and 2-choice were slower when presented first than when presented last. (III) Anomalies were responded to faster than literals and metaphors, except for literals when 2-choice preceded 3-choice. In that case, literals were faster than anomalies. (IV) Literals tended to be faster than metaphors (cf. main effect of expression type). However, this difference was not significant. (V) Condition effects were only significant for subjects as the random factor. (VI) No effects of preposition were significant.

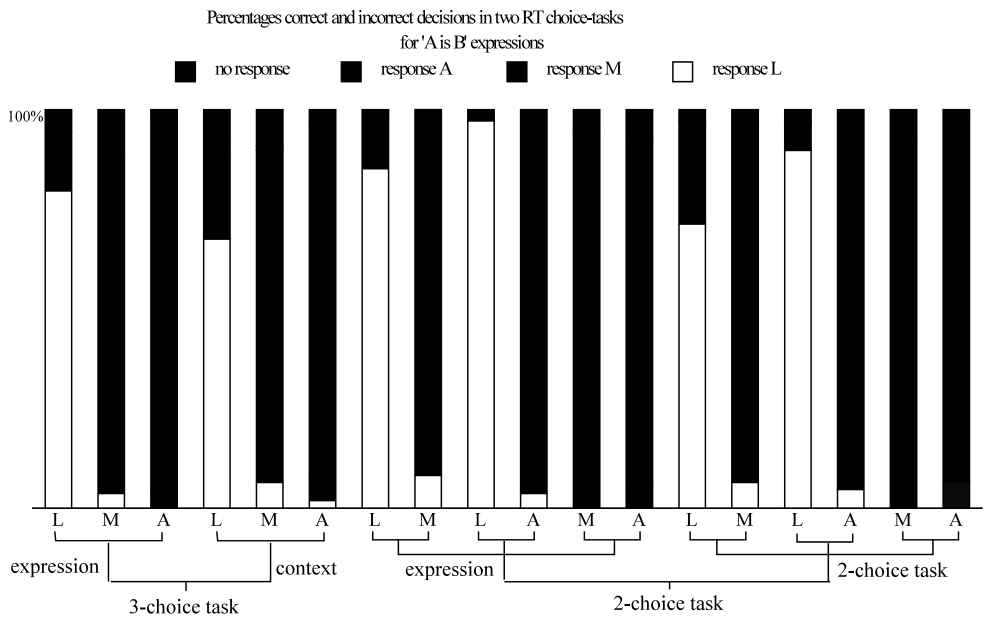
Figure 6.1: Grand mean arrival times (ms) for the interaction of task order by choice task by expression type (literal expressions (■) vs metaphors (▲) vs anomalies (●)). Results are pooled over the insignificant factors: Preposition, condition order, and condition. Bold lines denote 3-choice before 2-choice task, thin lines 2-choice before 3-choice.



Results for decision accuracy

Figure 6.2 shows the proportions of correct and incorrect decisions per choice task, condition and expression type.

Figure 6.2: Percentages of (in)correct decisions for expressions with or without 'like' in two choice tasks.



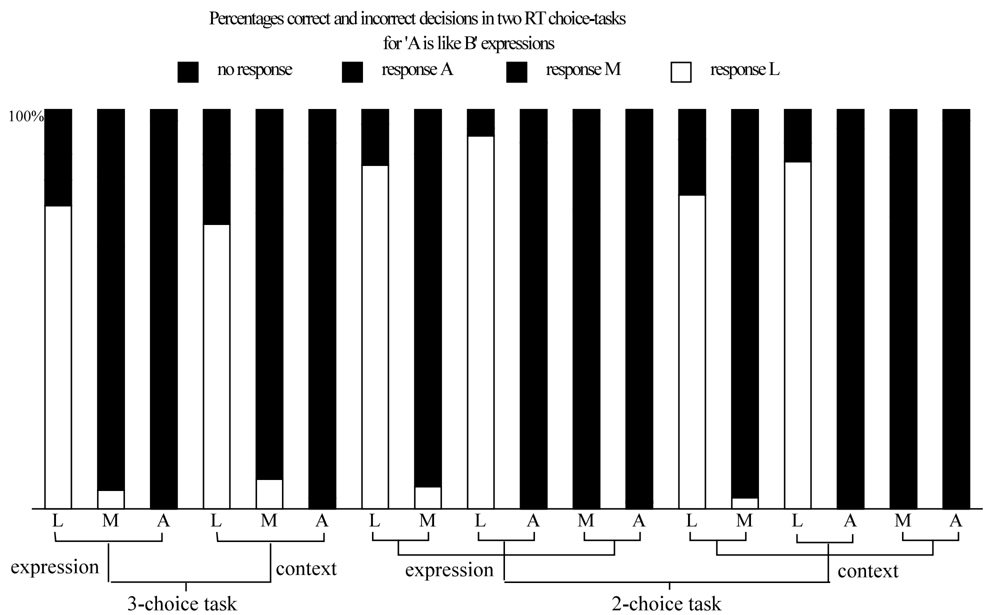


Table 6.1 (Appendix) provides the measure of agreement of the subjects for the correct decisions.

(I) For '*A* is *B*' constructions (upper panel), the 3-choice task in *context* rendered the lowest percentages correct responses and the 2-choice task in *expression* the highest. Correct responses ranged from 67.6%-97.3% for literals, 83.4%-99.2% for metaphors and 88%-96.4% for anomalies. (II) Most errors occurred for literals judged as metaphors in the 2-choice task in *context* (25.8%). The lowest number of errors occurred for anomalies seen as literals, ranging from zero to a maximum of 4.6% in the 2-choice task in *context*. (III) For all expression types of the form '*A* is like *B*' (lower panel), more errors occurred in the 3-choice than in the 2-choice task. For literals, the best and worst results were obtained in *context*. Metaphors scored lowest in *expression* and highest in *context*, whereas for anomalies, the reverse was valid. Correct decisions for literals ranged from 71.4%-93.6%, for metaphors from 61.2%-85.2% and for anomalies from 88.9%-99.1%. (IV) The highest error rates were found for metaphors judged as anomalies in the 2-choice task in *expression* (22.2%, lower panel). The highest proportion of literals judged as metaphors were found in the 3-choice task in *context* (14.8%). In none of the conditions, anomalies were judged as literals (0%). (V) Table 6.1 shows that the best results were found in the 2-choice task for '*A* is *B*' constructions (*expression*: κ = .908, *context*: κ = .828), whereas the most errors occurred in the 3-choice task for '*A* is like *B*' (*expression*: κ = .698, *context*: κ = .665).

Discussion of arrival times

An important barrier for evaluating the metaphor models was the strong order effect. Although the interaction of task order by condition order by choice task by condition was only significant for subjects as the random factor, its discussion may yet be elucidating. In the top panel of Figure 6.0, responses were fast when presentation was repeated (except for one case). Let 3-choice be abbreviated as 3, 2-choice as 2, *expression* as *e*, *context* as *c*, 'before' as -, and let I, II, III, and IV represent the first up to fourth presentation. Then, the interaction of task order by condition order by choice task by condition showed the following pattern for '*A* is *B*' constructions (grand mean RT_as are subsequent):

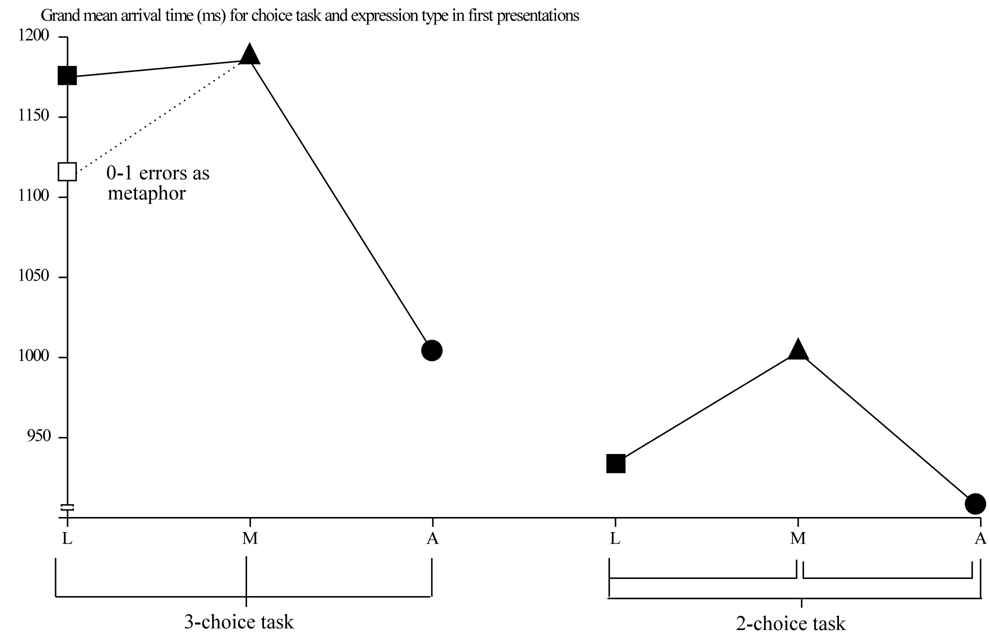
		I II III IV		I II III IV
bold solid	(3-2, *e*-*c*):	$3e > 3c > 2e > 2c$	RT_a :	$1093 > 984 > 853 > 813$
bold dash	(3-2, *c*-*e*):	$3c > 3e > 2c > 2e$	RT_a :	$1057 > 929 > 800 > 720$
thin solid	(2-3, *e*-*c*):	$2e > 2c > 3e < 3c$	RT_a :	$894 > 803 > 718 < 728$
thin dash	(2-3, *c*-*e*):	$2c > 2e < 3c > 3e$	RT_a :	$971 > 817 < 831 > 777$

Put differently, RT_a decreased as a function of repeated presentation of the stimuli for each task manipulation. The exception is the 3-choice task in *context* when presented last, in which case RT_a increased (bold signs). Although less clear, these patterns were found also for '*A* is like *B*' constructions. Figure 6.1 shows that this repetition effect was stronger in the 3-choice than in the 2-choice task (bold vs thin).

Thus, **differences between expression types were overthrown by order. As a serious methodological consideration for future research, then, expressions should not be repeated within subjects! Consequently, tests of the metaphor models are limited to the first reading of an expression**.

To inspect the effects of expression type in the first presentation, another MANOVA was performed with a design that deleted the factors of task order and condition order. Instead, only those values were analyzed that were obtained during the first presentation of an expression (cf. Figure 6.0: 3*e* bold solid, 3*c* bold dash, 2*e* thin solid, 2*c* thin dash). The only significant effects were the interaction between choice task and expression type, and the main effect of expression type (Appendix Table 6.2). The interaction is exemplified by Figure 6.3.

Figure 6.3: Grand mean arrival times (ms) for the interaction of choice task by expression type (literal expressions (■) vs metaphors (▲) vs anomalies (●)) during the first presentation of stimuli (correct responses only). Results are pooled over the insignificant factors: Preposition and condition. □ indicates the grand mean RT_a for those literal expressions in the 3-choice task that were judged as metaphors no more than once.



It may readily be seen that literal expressions and metaphors did not differ much in the 3-choice task. However, this result may be obscured by the following phenomenon. As shown by Figure 6.2, literal expressions were often erroneously judged as metaphors, which may indicate two things: Subjects sometimes made incorrect responses, or they occasionally perceived a literal expression as 'metaphorical'. Incorrect reactions should be recognizable by fast responses (speed-accuracy trade-off). Errors arising from doubtful cases, on the other hand, should slow down the grand mean of literal expressions. Indeed, RT_a for literal expressions that were judged as metaphors were slower than for correct decisions, sometimes more than 100 ms. Therefore, the incorrects were probably not due to speed-accuracy trade-off. Moreover, literal expressions that rendered few 'metaphor-errors' were faster than those with many. Figure 6.3 also shows the grand mean RT_a for literal expressions that were judged no more than once as metaphors (white block). Following this correction, the patterns for literals (white block), metaphors and anomalies in the 3- and 2-choice task become quite similar.

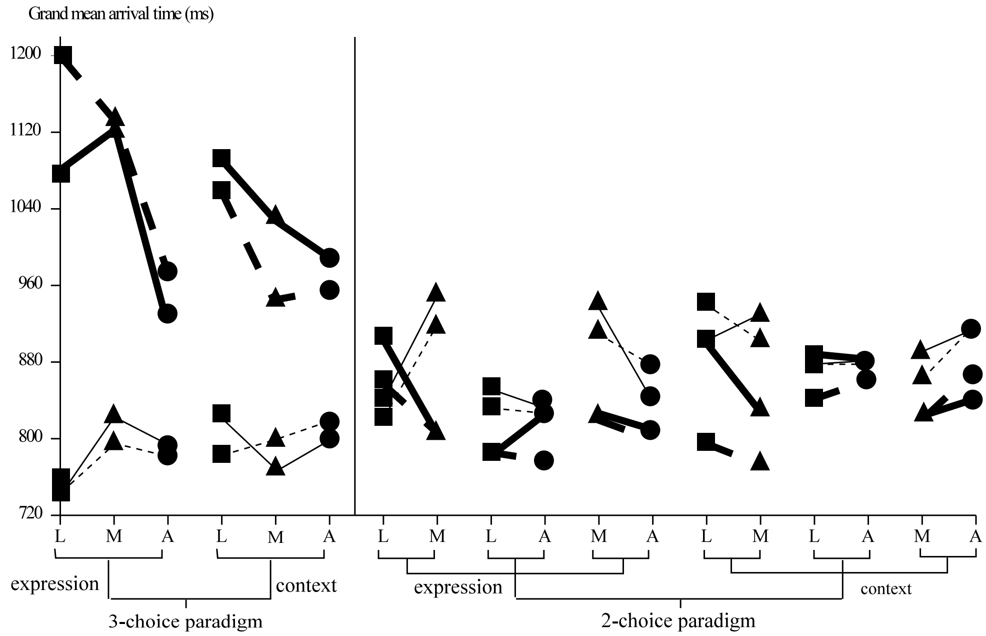
In sum, Figure 6.3 shows that RT_a was fastest for anomalies, and did usually not differ between literals and metaphors. Literal expressions that were hardly ever interpreted as metaphors differed most from real metaphors. No differences were found between *expression* and *context*. However, choice task affected RT_a most. In 2-choice, the overall response time was faster, while literals and anomalies differed less than in 3-choice.

None of the metaphor models are confirmed by these results, because none of them predicts that anomalies are faster than literals and metaphors. Although literals inclined to be faster than metaphors, this effect did not reach significance. According to the models, such a difference should have been found. Thus, **none of the metaphor models did well in accounting for the time course of metaphor processing**.

Check on word frequency

Low word frequencies of the anomalous *B-terms* might have been responsible for the great velocity of anomalies. Therefore - analogous to *Check 3* on word frequency in the previous Chapter - an extra level of high and low word frequency within *B-term* types was deviced as a within-subject factor. MANOVA was run on the grand mean arrival times for correct decisions in this 2*2*2*2*2*3*2 design of preposition, task order, condition order, choice task, condition, expression type and word frequency (high vs low). Figure 6.4 displays the grand mean arrival times, and Table 6.3 (Appendix) the relevant statistics.

Figure 6.4: Grand mean arrival times (ms) for task order, choice task, condition, expression type and word frequency. Results are pooled over the insignificant factors: Preposition and condition order. Bold lines indicate 3-choice before 2-choice task, thin lines 2-choice before 3-choice. Solid lines are high-, dashed lines low-frequent *B-terms*. ■ = literal expressions, ▲ = metaphors, ● = anomalies.



The results of Table 6.3 (Appendix) allow the following deductions:

(I) For 3-choice before 2-choice, high-frequent *B-terms* were slower than low-frequent ones in *context*, whereas in *expression*, the effect was the opposite. However, for 2-choice before 3-choice, the differences in *context* disappeared, whereas in *expression*, they reversed (interaction of task order by condition by word frequency). (II) In both the 3- and 2-choice task, high-frequent metaphoric *B-terms* were slower than low-frequent ones. For literal expressions, however, high-frequent *B-terms* were faster than low-frequent terms in the 3-choice task, and slower in the 2-choice task (interaction of task by expression type by frequency). (III) No interaction of anomalous *B-terms* with word frequency was significant.

In conclusion, **the low-frequent anomalous *B-terms* did not interfere with processing speed, and when word frequency affected RT_a, the effects were unsystematic. Word frequency was not more than a noise factor**.

Check on lexical ambiguity

Table 6.4: Grand means for arrival times (RT_a) of nonambiguous (n) vs ambiguous (y) *B-terms*. 3-2 = 3-choice before 2-choice, 2-3 = 2-choice before 3-choice, *e-c* = *expression* before *context*, *c-e* = *context* before *expression*.

Grand mean arrival time (ms) for (non)ambiguous terms

expression

	3-choice task						2-choice task											
	L_n	L_y	M_n	M_y	A_n	A_y	L_n	L_y	M_n	M_y	L_n	L_y	A_n	A_y	M_n	M_y	A_n	A_y
A is *B*																		
3-2																		
e-c	1008	985	1269	1124	1048	930	847	907	876	865	842	755	926	816	888	899	890	837
c-e	1029	1073	991	956	858	803	743	958	687	759	652	697	673	676	712	761	786	738
2-3																		
e-c	676	659	729	781	752	736	884	848	896	1104	897	866	818	792	966	961	860	843
c-e	709	757	810	798	791	773	784	773	884	861	809	895	845	770	794	864	843	825
A like *B*																		
3-2																		
e-c	1258	1396	1064	1353	970	1035	875	883	1013	992	875	766	841	832	948	1028	772	794
c-e	1053	956	1041	1025	891	961	832	763	965	986	838	774	728	828	894	730	862	752
2-3																		
e-c	810	770	984	908	789	816	854	844	939	958	999	892	906	918	1024	995	902	922
c-e	847	797	896	973	829	820	845	816	949	908	750	788	820	791	878	904	825	805

context

	3-choice task						2-choice task											
	L_n	L_y	M_n	M_y	A_n	A_y	L_n	L_y	M_n	M_y	L_n	L_y	A_n	A_y	M_n	M_y	A_n	A_y
A is *B*																		
3-2																		
e-c	936	995	965	993	932	1002	801	960	839	732	822	785	826	845	772	699	804	859
c-e	1205	1017	1107	947	1152	965	789	768	729	857	860	899	846	770	737	727	791	782
2-3																		
e-c	690	742	733	682	764	733	754	812	800	822	852	750	775	780	740	845	899	897
c-e	843	833	818	880	835	817	1005	1166	916	1149	954	1041	906	949	869	915	874	920
A like *B*																		
3-2																		
e-c	1156	1132	1197	1021	1020	1045	867	815	783	832	841	843	899	970	942	861	888	899
c-e	1171	1106	1127	1065	1031	1114	920	872	920	843	831	880	875	825	790	995	1103	856
2-3																		
e-c	739	752	631	669	808	807	812	895	808	973	912	758	943	875	720	920	920	1015
c-e	883	860	935	806	876	866	1015	1018	1016	989	924	901	952	907	1121	1120	910	895

Lexical ambiguity may affect RT, because the availability of multiple meanings may lead to a conflict in selecting among the options. Thus, the RT for ambiguous terms may be elongated in comparison with nonambiguous terms.

To test the effects of lexical ambiguity, an extra within-subject factor of lexical ambiguity of the *B-term* was included in the analysis. MANOVA was run on the grand mean arrival times for correct decisions in a 2*2*2*2*2*3*2 design of preposition, task order, condition order, choice task, condition, expression type and lexical ambiguity (no vs yes). Table 6.4 shows the grand means and Table 6.5 (Appendix) the statistics.

The results of Table 6.5 show the following:

(I) For all significant interactions, lexical ambiguity increased and decreased RT_a about as often. (II) For literals and metaphors in the 3-choice task, ambiguous terms were faster than nonambiguous terms. In the 2-choice task, no difference was found for literals, whereas the effect for metaphors reversed, making ambiguous terms slower. In comparison with metaphors, anomalies showed that ambiguous terms were faster than nonambiguous terms, which increased from the 3-choice to the 2-choice task (liberal α-level).

Thus, **ambiguity affected RT_a in an inconsistent way, so that it may be considered noise**.

Discussion of decision accuracy

Although the expression categorization experiment in Chapter 4, Table 4.1 suggested a strong consensus on the status of the expressions, the decisions presented in Figure 6.2 showed otherwise. It is well-known that judgements may be unanimous, whereas the actual decisions in a forced decision RT paradigm show hesitation and a nonnegligible percentage of errors (Sanders, in press). This was particularly the case with literal expressions that were occasionally perceived as metaphors. In these cases, RT_a for the errors was longer than for the correct choices, which is usually the case when stimuli can be discriminated poorly (Sanders, in press). In other words, the errors were not so much due to speed-accuracy trade-off, as to a poor discrimination of the stimuli.

This may be evidence for a continuum between literal expressions and metaphors, with an overlapping 'twilight' zone in between. It should follow that metaphors which are occasionally interpreted as literal expressions are also slower than clear metaphors, although the decision data show that this type of error is seldom made. The longer reaction times for literal expressions with a metaphoric bias may also suggest that the criterion for deciding between the two has no fixed value. Instead, it may be an area with a lower and upper limit, leading to hesitations and long response times.

Nonetheless, the error rates were high. It may be inferred that the nine prototypical expressions used in the 90 training trials were not diverse enough to distinguish the expression type of new instances. Therefore, it may be advisable to train subjects first with a limited set of repeated prototypes to stipulate the different expression types, and then to train them with a larger number of unrepeated (unique) expressions - not related to the experimental set - until the errors are less than 10% of the total number of responses (Sanders 1980).

6.5 General discussion for psychology

Could the metaphor models be adjusted post hoc in order to obtain correspondence with the RT_a data? **Comparison models (either serial or parallel) cannot adequately account for the fast anomalies in combination with the about equal processing times for literals and metaphors**. The first point - fast anomalies - is not fatal if the order of checking the criteria is inverted. The first criterion checks whether too many shared features are found to call the expression an anomaly. Then, metaphors and literal expressions are differentiated by a second criterion, asking whether too many shared features are found to decide for metaphor. In that case, processing speed is regulated by the order of checking the criteria (fast anomalies, equal literals and metaphors), while the decision itself is based on the shared set size.

However, the results of weighted shared set size (Chapter 5) showed that literal expressions and metaphors did not have different shared set sizes. This means, that subjects should not have been able to decide between literal expressions and metaphors. Although, indeed, most of the mix-up occurred between these two expression types, broadly speaking, the corresponding decision accuracy was between 60% and 95%.

A principal problem for the interaction model (either serial or parallel) is that the effects of context appear to be ineffective. In all cases, differences between *expression* and *context* were insignificant. Again, fast anomalies can be explained by reversing the order of checking the response criteria. In the first stage, anomalies are judged on the shared set size of normal features. If there are too many shared features, the expression is not an anomaly, and a second criterion is checked for literal expressions. If the shared set size is not large enough to make the expression literal, a second stage is commenced, in which sufficient fitting relations make the expression a metaphor. However, in that case, metaphors should have taken much more time to process than literal expressions, which was not the case - although a trend was present.

Equal processing times can only be explained by a parallel model, in which the establishment of shared normal features and shared relations is equally fast. The distinction between literal expressions and metaphors, then, would be based on the type of information that is accumulated: Shared normal features for literals, shared relations for metaphors. On the other hand, the feature elicitation results showed no significant differences among expression types for the accumulation of any form of relation overlap.

No serial anomaly model can explain the fast anomalies in combination with the approximately equal processing times for literals and metaphors. Again, the latter point is crucial, because literal overlap was supposed to be checked before figurative overlap (or perhaps reversely), which should have been reflected in different processing times. However, **a parallel anomaly model perfectly explains the results of feature elicitation and reaction times**. In the feature elicitation experiments, it was found that anomalies yielded the smallest shared sets, either literal or figurative. Literal expressions established large literal shared sets and small figurative ones. Metaphors established equally large literal shared sets. However, they also established large shared sets of features that were literal for the one term and figurative for the other (S_{lf} and/or S_{fl}). Taken in unison these results and the reaction time pattern that anomalies were fastest - whereas literal expressions were only insignificantly faster than metaphors - the following axioms would underlie a parallel anomaly model:

1. There are two parallel accumulators: The literal shared set (S_{ll}) and the figurative shared set (S_{lf} and/or S_{fl})
2. The shared set is completed before response execution
3. The criterion q_l for 'literal' is checked simultaneously with q_f for 'metaphor'

Since q_l and q_f can be checked at the same time, expression type differences are due to a race between the two accumulators S_{ll} and (S_{lf} and/or S_{fl}). Thus, response speed is not limited by the *response phase*, but rather by the parallel *activation phases* and parallel *comparison phases*.

Expressions with small feature sets and little overlap are processed most quickly, so that anomalies are rejected almost instantly by q_l and q_f at the same time. Literal expressions accumulate mainly literal overlap (S_{ll}), whereas the small figurative overlap is unimportant. Metaphors accumulate literal (S_{ll}) as well as figurative overlap (S_{lf} and/or S_{fl}). Literal expressions may be reacted to faster than metaphors, when the figurative sets are exhausted quickly, and time is only spent on establishing the shared literal set. Metaphors may be slower, when the accumulation of two sorts of information (literal and figurative) takes more time than accumulating one kind (literal).

In other words, anomalies are processed in a short literal and a short figurative stage. Literal expressions are processed in a long literal stage and a short figurative one, and metaphors in a long literal and a long figurative stage. This model explains the low RT_as for anomalies, and the tendency for literals to be faster than metaphors. It is also possible that the literal expressions and metaphors tie. In that case, literal and figurative information is accumulated equally fast, and accumulating two sources of information is not more time consuming than accumulating one source. Consequently, the categorization of expressions is based purely on different types of information. When metaphors yet are slower, it may be due to the accumulation of two information types, or that the activation of figurative features starts a little later, as illustrated in the previous Chapter (Figure 5.15, Chapter 5).

It should be recognized, however, that this account only applies to the first presentation of expressions. At repeated presentation, all differences tend to deteriorate. Subjects probably recognize an earlier judgement, so that the matching of features becomes unnecessary. Furthermore, differences between literals and metaphors may have been drowned by noise factors such as word frequency and lexical ambiguity.

With regard to the next Chapter, the serial anomaly model of Figure 2.1, Chapter 2, envisioned a covert effect of 'conclude anomaly' when the criterion q_l for literal expressions was not met. This anomalous moment would be the catalyst of the second, figurative, stage. Chapter 7 investigates whether this covert effect is materialized in the evoked potential N400. Moreover, if the parallel rather than the serial anomaly model is operative, can the N400 be considered a part of the process, is it a surprise effect, or a supplementary after effect?

6.6 General discussion for the theory of literature

What could reaction time say about literature? Not very much, as long as the interference of preliminary variables is little-known. The RT results showed that the method of experimentation had strong effects, from which the factors of interest (preposition, condition, expression type) suffered a lot.

The effects of preposition and condition were negligible, indicating that metaphor and simile constructions were not processed differently, and that the poetic context did not help to identify an expression type. The preposition *like* may cue the interpretation into the direction of metaphor; however, the (in)appropriate combination of terms seemed more important when making a correct judgement. The interjection of the preposition is probably a subordinate aid. Since *context* did not yield important differences in comparison with *expression*, it may be inferred that expressions are judged on sentence level, without relying too much on the poetic context. This result agrees with the feature elicitation, where differences between *expression* and *context* usually were insignificant.

As far as expression type is concerned, it was found that anomalies were processed the most quickly, whereas literal expression and metaphors hardly differed. Literals tended to be faster than metaphors, although this effect was not statistically reliable. That literals and metaphors were processed equally fast was also found by Glucksberg, Gildea & Bookin (1982), Gerrig & Healy (1983), Gildea & Glucksberg (1983), Inhoff, Lima & Carroll (1984) and Gerrig (1989_b). However, it should be noted that in the present experiments, these differences among expression types were a **limited** result, because they were sensitive to choice task (3-choice vs 2-choice) and task order (3-choice before 2-choice and vice versa). Moreover, when the expressions were repeated, the differences among expression types faded out. In other words, the range of situations that a metaphor model can cover is very restricted, unless it accounts for various effects of task manipulation.

From these observations, it follows that the earlier RT experiments in psychology (see also the previous paragraph) are strongly circumscribed by the experimental set up. Moreover, Section 6.2 demonstrated that almost all designs were liable to substantial flaws and were usually analyzed improperly.

With respect to the time course of metaphor processing, then, we may start from scratch. Before all else, the expressions should be matched on the relevant variables (cf. Chapter 4) even better than in the presently-used stimulus set. Lexical ambiguity and word frequency may conjure too much noise, while repetitions of stimuli overshadow the effects of interest. Also the kind of choice task should be considered as a forceful factor. Obviously, readers process expression types differently, when they know beforehand that all options are available (literals, metaphors, anomalies) or that the choice is narrowed down to, for example, metaphors and anomalies.

The last remark may have interesting consequences for the effects of genre. When paper articles are contrasted with poetry, it may be that in the journalistic text, readers expect fewer metaphors than literals, and probably no anomalies. In the poem, however, they may expect more metaphors and perhaps even more anomalies than literal expressions. In other words, reading papers is a 2-choice task between literals and metaphors, with a higher probability for the first expression type (no counterbalance). Reading poetry is a 3-choice task among metaphors, anomalies and literals, with a corresponding decrease in probability (no counterbalance). Differences between the two genres, then, may be explained not only by the mix of more and more difficult expression types in poetry, but also by the fact that to perform a 3-choice task simply is a tougher job than to do a 2-choice task.

Given the limitations outlined above, Section 6.5 revised the serial anomaly model as a **parallel anomaly model**. Taking together the results of feature elicitation and reaction times, this model has the most explanatory power. It will be recalled that the feature elicitation results indicated that anomalies found the smallest literal and figurative shared sets, that literal expressions found large literal shared sets and small figurative ones, whereas metaphors found large literal as well as figurative shared sets. Note that the figurative shared sets were composed of features that were literal for the one term and figurative for the other. The reaction times showed that anomalies were fastest, whereas literals and metaphors were processed equally fast, with a slight advantage for literals. The parallel anomaly model postulates that:

1. There are two parallel accumulators: The literal shared set (S_{ll}) and the figurative shared set (S_{lf} and/or S_{fl})
2. The shared set is completed before response execution
3. The criterion q_l for 'literal' is checked simultaneously with q_f for 'metaphor'

The threshold criteria q_l and q_f are checked synchronously, so that expression type differences are the result of a race between the two accumulators S_{ll} and (S_{lf} and/or S_{fl}). Thus, the *activation phases* for literal and figurative features are working in parallel, just as the two *comparison phases*.

The feature elicitation results showed that anomalies established small shared literal and figurative sets. Thus, activation and comparison finish quickly, and the anomaly is rejected simultaneously by q_l and q_f as literal or metaphor. Literal expressions accumulate literal overlap (S_{ll}), whereas the figurative overlap remains small. Metaphors accumulate literal (S_{ll}) as well as figurative overlap (S_{lf} and/or S_{fl}). Thus, literal expressions may be faster than metaphors, when they only establish literal shared sets, whereas the figurative shared set is so small that it does not take much time. Metaphors may be slower, when they accumulate a literal as well as a figurative shared set.

Thus, anomalies are processed in two short stages, literals in a long literal stage and a short figurative stage, and metaphors in a long literal and a long figurative stage. If q_l and q_f are checked in parallel, it is also possible that no differences in processing speed are found between literal expressions and metaphors. In that case, literal and figurative information is accumulated equally fast. Although metaphors have equally large shared literal sets, they are not categorized as 'literal', because they also exceed q_f . Metaphors may be slower when the accumulation of two information types takes longer than the accumulation of one type, or when the activation of figurative features begins later in the associative flow, as represented by Figure 5.15, Chapter 5.

In the serial anomaly model, the transition from the literal to the figurative stage was accompanied by a 'shock at the clash of meanings' (Beardsley 1982: 267). This shock was supposed to be an inalienable part of the process, and was supposed to be an incentive to advance into the second stage. The next Chapter surveys whether this shock can be measured inside the brain itself. The psychophysiological literature suggests that certain potentials in the brain are sensitive to semantic deviations. Since the serial anomaly model does not seem as powerful as the parallel version, the status of the electrocortical shock as part of the metaphor process is questioned. Is it the impetus to enter the second stage, merely an additional surprise effect, and if not, what is it?

Notes:

1. An extra level of no-context would be even better.

2. Actually, this implies two anomaly types, distinguished by smaller or larger literal shared sets.

Appendix to Chapter 6

Table 6.0: MANOVA for the effects of preposition, task order, condition order, choice task, condition and expression type on the grand mean arrival time (RT_a) for correct responses. Only effects at least significant for subjects as the random factor are tabulated.

Interaction of task order by condition order by choice task by condition:

dep. variable	$F_{1,16}$	p	df	quasi-F	p
RT_a	10.69	.005	$_{6,2}$	.02	>.50

Interaction of task order by choice task:

1 = ((3 before 2) vs (2 before 3)) vs ((3-choice) vs (2-choice))

dep. variable	$F_{1,16}$	p	df	quasi-F	p
RT_a	196.45	.000	$_{20,2}$	.05	>.50

Interaction of condition order by condition:

1 = ((*expression* before *context*) vs (*context* before *expression*)) vs (*expression* vs *context*)

dep. variable	$F_{1,16}$	p	df	quasi-F	p
RT_a	19.00	.000	$_{2,2}$	.22	>.50

Interaction of task order by choice task by expression type:

2 = ((3 before 2) vs (2 before 3)) vs ((3-choice) vs (2-choice)) vs (literals vs anomalies)
3 = ((3 before 2) vs (2 before 3)) vs ((3-choice) vs (2-choice)) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{3,14}$	p	df	quasi-F	p		coeff.	t	p
RT_a	.51	4.93	.015	$_{3,7}$	7.15	.015	2	177.48	3.80	.002**
							3	140.60	3.03	.008**

Interaction of task order by expression type:

2 = ((3 before 2) vs (2 before 3)) vs (literals vs anomalies)

dep. variable	Pillai's Trace	$F_{2,15}$	p	df	quasi-F	p		coeff.	t	p
RT_a	.39	4.83	.024	$_{2,33}$	46.68	.000	2	181.54	3.10	.007**

Interaction of condition order by condition by expression type:

dep. variable	Pillai's Trace	$F_{2,15}$	p	df	quasi-F	p
RT_a	.52	8.13	.004	$_{2,4}$	.31	>.50

Interaction of condition by expression type:

dep. variable	Pillai's Trace	$F_{2,15}$	p	df	quasi-F	p
RT_a	.67	15.36	.000	$_{2,4}$	.33	>.50

Main effects of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,15}$	p	df	quasi-F	p		coeff.	t	p
RT_a	.48	7.06	.007	$_{2,4}$	33.88	.009	1	-90.56	-1.49	.155
							2	111.78	1.91	.074
							3	202.34	3.86	.001**

Table 6.1: Cohen's Kappa for the correct decisions of each subject in two choice tasks for expressions with or without *like* in *expression* and *context*.

	A* is *B					***A* is like *B***			
	3-choice		**2-choice**			**3-choice**		**2-choice**	
	expr.	***context***	***expr.***	***context***		***expr.***	***context***	***expr.***	***context***
subject	**κ**	**κ**	**κ**	**κ**	**subject**	**κ**	**κ**	**κ**	**κ**
01	.89	.60	.86	.86	13	.45	.63	.86	.81
02	.56	.22	.80	.69	14	.46	.57	.80	.86
03	.54	.47	.94	.86	15	.27	.46	.67	.78
04	.72	.78	.86	.92	16	.79	.50	.83	.87
05	.51	.39	.94	.74	17	.29	.27	.77	.69
06	.94	.68	1.00	.86	18	1.00	.74	.97	1.00
07	.95	.94	.94	.94	19	.79	.73	.69	.69
08	1.00	.94	.83	.81	20	.83	.67	.81	.94
09	.89	.78	.92	.78	21	.67	.73	.83	.68
10	.94	1.00	.97	.86	22	.94	.89	.92	.78
11	.78	.83	.92	.86	23	.89	.79	.94	.81
12	.88	.89	.92	.75	24	1.00	1.00	1.00	.69
$\bar{x}$	**.800**	**.710**	**.908**	**.828**		**.698**	**.665**	**.840**	**.800**

Table 6.2: MANOVA for the effects of preposition, choice task, condition and expression type on the grand mean arrival time (RT_a) in the first presentation of stimuli. Only significant effects are tabulated.

Interaction of choice task by expression type:

1 = ((3-choice) vs (2-choice)) vs (literals vs metaphors)
2 = ((3-choice) vs (2-choice)) vs (literals vs anomalies)
3 = ((3-choice) vs (2-choice)) vs (metaphors vs anomalies)

dep. variable	Pillai's Trace	$F_{2,15}$	p	df	quasi-F	p		coeff.	t	p
RT_a	.48	6.87	.008	$_{2,4}$	34.27	.008	1	31.55	.87	.396
							2	71.37	3.16	.006**
							3	39.82	1.61	.128

Main effects of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,15}$	p	df	quasi-F	p		coeff.	t	p
RT_a	.80	29.76	.000	$_{2,4}$	35.31	.005	1	-39.83	-1.10	.287
							2	102.05	4.52	.000**
							3	66.62	2.68	.016**

Table 6.3: MANOVA for arrival times (RT_a) of high-frequent vs low-frequent *B-terms*. Only significant effects are listed.

Interaction of condition by word frequency:

1 = (*expression* vs *context*) vs (high vs low)

dep. variable	$F_{1,16}$	p		coeff.	t	p
RT_a	6.01	.026	1	-355.72	-2.45	.026*

Interaction of task order by condition by word frequency:

1 = (3-2 vs 2-3) vs (*expression* vs *context*) vs (high vs low)

dep. variable	$F_{1,16}$	p		coeff.	t	p
RT_a	9.73	.007	1	-452.55	-3.12	.007*

Interaction of task by expression type by frequency:

1 = (3-choice vs 2-choice) vs (literals vs metaphors) vs (high vs low)

dep. variable	Pillai's Trace	$F_{3,14}$	p		coeff.	t	p
RT_a	.99	526.99	.000	1	-3492.48	-42.00	.000**

Table 6.5: MANOVA for the grand mean arrival times (RT_a) of nonambiguous (n) vs ambiguous (y) *B-terms*. Only significant effects are listed.

Interaction of preposition by task order by condition order by condition by ambiguity:

1 = (with like vs without like) vs ((3 before 2) vs (2 before 3)) vs ((*expression* before *context*) vs (*context* before *expression*)) vs (*expression* vs *context*) vs (amb_n vs amb_y)

dep. variable	$F_{1,16}$	p		coeff.	t	p
RT_a	15.27	.001	1	665.93	3.91	.001*

Interaction of preposition by condition order by task by condition by ambiguity:

1 = (with like vs without like) vs ((*expression* before *context*) vs (*context* before *expression*)) vs (3-choice vs 2-choice) vs (*expression* vs *context*) vs (amb_n vs amb_y)

dep. variable	$F_{1,16}$	p		coeff.	t	p
RT_a	4.88	.042	1	396.18	2.21	.042*

Interaction of task by expression type by ambiguity:

1 = (3-choice vs 2-choice) vs (literals vs metaphors) vs (amb_n vs amb_y)
2 = (3-choice vs 2-choice) vs (literals vs anomalies) vs (amb_n vs amb_y)
3 = (3-choice vs 2-choice) vs (metaphors vs anomalies) vs (amb_n vs amb_y)

dep. variable	Pillai's Trace	$F_{3,14}$	p		coeff.	t	p
RT_a	.99	362.94	.000	1	-3557.61	-33.70	.000**
				2	-8.93	-.15	.880
				3	131.24	2.23	.040✔

✔ not significant at $\alpha = .0166$

CHAPTER 7: ELECTROENCEPHALOGRAPHY*

7.0 Psychophysiological traces of metaphor

One of the exciting endeavors in psychology is to explain functional models from physiology. The anomaly model in particular employs a physical incident as the marker of extra cognitive processing. In the theory of literature, this 'shock' of metaphor can be seen as an indicator of foregrounding. In Section 7.1, the sparse examples of physiologically oriented ideas on literary reception are attended, and Berlyne's arousal theory for artistic stimuli is discussed. In Section 7.2, studies are reviewed on the linguistically interesting N400-effect in the electroencephalogram (EEG). This N400 - a potential supposedly susceptible to semantic anomalies - is opted for by Section 7.3 to experimentally investigate the 'shock' effect of metaphor in Section 7.4. Section 7.5 and 7.6 discuss the importance of the N400 for metaphor research in psychology and the theory of literature.

7.1 Theory of literature: Prudent allusions at psychophysiology

Intuitively, the connoisseur of literature is reluctant to accept that something so subtle as poetry is done justice to by apparatus such as voltmeters and electrodes, and rightly so. Personal heuristics can never be replaced by measuring electric brain activity at the scalp to reveal what the poem means. On the contrary, using the tangle of amplitude shifts that arises from such measurements for the interpretation of a text is a worse case of soothsaying than the empiricists blame the hermeneutics for. In addition, it would be far easier to ask people what the poem means than to measure their brain signals while reading. Else, nature would have supplied us with electrodes instead of a tongue.

What personal heuristics cannot do, however, is to give a controllable account of what happens in the brain. When a scholar of literature asserts that certain text elements are put at unexpected positions in the clause - 'scintillating this poem with wit' - we have to take this claim for granted. It remains unclear whether a real reading experience was described, or whether the statement came from a learned book on stylistics. In other words, the empirical investigator of literature will be reluctant to believe that the scholar's claim was not a result of consensus, entailed by the academic environment rather than the text itself. In addition, surprise effects may be so subtle that personal heuristics is too coarse a measure as to access them.

---

* Notes are on page 304. Appendix with statistics on the pages 305-310.

In this respect, an often made mistake is that introspection would not be an indirect accessment of knowledge, even about oneself. In *Politeia*, Plato stressed that the complex inner world can only be captured through imperfect representations, the puppet players' shadows on the wall. Neurophysiological measures are no different from introspection for that matter (Kok 1988: 3), except that they are sensitive to text effects at the micro-level of processing and that they provide a controllable means to study those effects.

Much of the unwillingness in scholars of literature may be explained by neuropsychology's reconciliation of two apparently opposite modes: Mind and matter. For the scholars of literature, the article of faith 'mind over matter' holds, whereas for neuropsychology the creed is that 'mind is matter'. Changeux (1985) indicated that as early as the Church Fathers (fourth and fifth century), this dichotomy dominated Western philosophy of the mind. Nemesius and Augustinus placed the functions defined by Aristotle such as ratio, perception, memory and imagination not inside the brain itself, but in its hollow ventricles (cf. Kok 1988: 5). Descartes thought that cognition was not spatial, whereas physiological processes were. This was perhaps the most influential idea for the persistent separation of mind and matter in the present day, although modern physics tried to avoid the issue in postulating that 'all is energy'.

Despite the possibility of a false distinction between mind and matter or of totally interacting energies, the fact remains that information is transmitted via physical entities such as sound (speaking, hearing), light (reading, seeing) and pressure (touching). Thus, if there is a rift between mental and brain processes, something must be bridging the gap in translating the physical impulses into mental meaning, and vice versa. In *Passions de l'âme*, Descartes also found this transition to be self-evident. Unless the transition is completely chaotic, certain physical regularities correlate with psychological phenomena, including reading poetry. Positively, the theory of literature sometimes meditated the potentiality of a 'biomechanics of language'.

> (...) speech is a universally human and exclusively human property imperatively call[ing] for an attentive inquiry into the biological prerequisites of human language. Bloomfield's reminder that among the special branches of science, linguistics "intervenes between biology, on the one hand, and ethnology, sociology, and psychology, on the other" (...) is most opportune. (Jakobson 1971: 675)

Jakobson specified his recommendation in explicitly stating that:

> This research becomes particularly productive when results of linguistic analysis are matched with the PSYCHOPHYSICAL data. (Jakobson 1971: 689)

Primary, Jakobson imagined the study of language impairments to fulfill the need of neurolinguistic knowledge:

> The deepest discernment of the relation between the human organism and its verbal abilities and activities is achieved by the mutual help of NEUROBIOLOGISTS and linguists in a comparative inquiry into the various lesions of the cortex and the resulting aphasic impairments. (...) when we confront this strictly linguistic framework with the anatomical data, it proves to coincide with the topography of the cerebral lesions responsible for the diverse impairments (...). The prospective development of such interdisciplinary, "neurolinguistic" research in aphasic and psychotic speech (...) will undoubtedly open new vistas for a comprehensive study of the brain and its functions as well as for the science of language and other semiotic systems. A deeper insight into the biologic foundations of language may be expected from the ongoing experience with split-brain operations. (Jakobson 1971: 688)

And indeed, studies on dyslexic and aphasic patients unravelled new insights into language processes, evolving in, for instance, the Wernicke-Geschwind model of reading (Benson 1981: 69-90, Hynd 1989: 123-148). Even with regard to metaphors, research was performed with brain-damaged subjects (Winner & Gardner 1977, next Section).

Whereas Jakobson suggested the cooperation between psychophysiology and linguistics, his pupil Abernathy connected the study of literature (the 'other semiotic systems') with the recording of EEGs. Abernathy speculated that the development of a dream described in a poem by the Russian poet Blok might be retraceable in the brain waves of sleep-wake states recorded by an EEG. First, he outlined that:

> (...) we happen to be unusually well-informed about the origins of this poem: its subtitle "A dream" is meant literally, since by Blok's own testimony it is based on an actual dream he experienced in November 1905 and on his written notes made a day and a half later, still under the influence of the powerfull dream-impression. This preliminary sketch of the poem survives, and comparison of it with the final published text makes plain that the latter is directly founded on this 'dream-stuff' only as far as line 249. Lines 250-304 lack such antecedents, and in their content reflect quite a different mood, dispelling the nightmare atmosphere of the swamp (...). It is reasonable to see in this a product of the artistic reworking, (...), of the original dream-nucleus. (...) from a formal analysis (...) this 'nucleus' appears as a recognizable structural whole, embedded in a framework - which, for the rest, reproduces the nuclear pattern in such a way that

> the same characterization of the total effect, as one of form gradually emerging from chaos. (Abernathy 1967: 8-9)

Abernathy analyzed that in the dream poem, form gradually emerged from chaos, and added in a footnote that this formal aspect might be represented in the EEG:

> Given the poem's dream-ancestry, it is tempting though probably fanciful to try to correlate this progressive regularization of its form more or less directly with features of hypnoidal (hypnopompic) states - say with the behavior of electroencephalograph tracings during gradual awakening, showing replacement of the slow irregular delta waves of deep sleep first by more or less random bursts of activity and finally by the waking alpha rhythm. (Abernathy 1967: 9)

Notwithstanding this nice metaphor on dream description and electrocortical effects of dreaming, EEGs can be exploited seriously to examine the theory of literature.

Mukařovský (1964: 17-30), for instance, suggested that although phonological and semantic deviations are not absent in standard language, their *degree* and *mode* are different than in literature. In certain respects, these observations were confirmed by Hoorn (1996). Particular components of the EEG - the so called event-related brain potentials - appeared to be sensitive to semantic and phonological deviations in reading four-lined verses in alternating rhyme. The last word of each verse either rhymed or not and was a semantically correct completion or not. Confirming Mukařovský, the brain potentials were smaller for rhyming and semantically correct completions (the more standard language) as opposed to nonrhyming and semantically deviant endings, which evoked higher amplitudes. Likewise, such differences in degree of deviateness - reflected in the height of ERP amplitudes - may be found also when literal expressions are compared with metaphors.

Here is where the theory of literature converges with the theory of psychophysiology. Both disciplines are interested in the effects of semantic unexpectedness and surprise. As indicated by the previous Chapter, Šklovskij (1965: 12) based the theory of foregrounding on the transgression of automatized language processing. Art made objects unfamiliar and difficult, a position subscribed also by Striedter:

> First, defamiliarization impedes the kind of perception automatized by linguistic and social conventions (...). Second, in a kind of countermovement, by impeding perception, defamiliarization directs perception to the estranging and impeding from itself. (Striedter 1989: 23-24; Zwaan 1993: 41)

Thus, what could be more promising than a tool that distinguishes impediment from satisfaction of expectations in language perception? One particular kind of event-related brain potential (ERP) seems most sensitive to semantic deviateness. Kutas & Hillyard (1980_a , 1980_b , 1981) presented words that completed sentences in diverse degrees of nonsensicality ('he took a sip from the transmitter', 'he mailed the letter without a check') as opposed to normal completions ('he took a sip from the coffee', 'he mailed the letter without a stamp'). During the presentation of the completing word, the EEG was sampled at various locations on the scalp. The EEG averaged over subjects showed that a negative polarity (N) was found, 400 milliseconds after presentation of the anomalous word. The height of the amplitude of this ERP labeled N400 varied systematically with the degree of anomality. In the next Section, a review of the N400-literature is provided, in which various stimulus types and manipulations are discussed for their relation with N400-elicitation, along with some critical notes on brain signal interpretation.

For the theory of literature, however, study of the N400 in response to literary stimuli would be a giant leap forward, since it would provide a hypersensitive utensil to investigate the *real effects* of stylistics, reviving the long neglected inheritance of the Russian Formalists and Prague Structuralist School.

To go straight to the point, investigating brain potentials could provide a solution to the conflict about whether and to which extent metaphoric creations are deviations or regular processes of certain stylistic varieties. For instance, Jakobson took the position that metaphors are no different from regular expressions.

> Metaphoric creations are not deviations but regular processes of certain stylistic varieties, which are subcodes of an overall code, and within such a subcode there is nothing deviant in Marvell's figural assignment of a concrete epithet to an abstract noun (properly a 'hypallage') - "a green Thought in a green shade" - or in Shakespeare's metaphoric transposition of an inanimate noun into the feminine class - "the morning opens her golden gates" - or in the metonymic use of "sorrow" instead of "sorrowful while", which Putnam's paper excerpts from Dylan Thomas ("A grief ago I saw him there"). In contradistinction to the agrammatical constructions as "girl sleeps" the quoted phrases are meaningful, and any meaningful sentence may be submitted to a truth test exactly in the same way as the statement "Peter is an old fox" could lead to a reply "It's not true; Peter is not a fox but a swine, while John is a fox." Incidentally, neither ellipses nor reticence or anacoluthon could be considered as deviant structures; they, and the slurred style of speech, a brachylogical subcode to which they belong, are merely lawful derivations from the kernel forms embedded in the explicit standard. (Jakobson 1961: 252)

Jakobson opposed the view that metaphors are deviant structures. They are not agrammatical and can be submitted to a truth test, so that they are meaningful expressions. Therefore, metaphor is not a deviant form of language use. Culler on the other hand, advanced that metaphor *is* deviant, not as a matter of structure, but as a matter of *effect*, owing to its false assertion of identity. In line with Jakobson, Culler subscribed that metaphoric structures are no different than literal expressions, not even in meaning. In advocating the philosopher Davidson, however, Culler accentuated the difference in effect:

> By denying that metaphors have any special meaning (they assert literal equivalences which are false) [Davidson] makes necessary an elaborate account of the *effect* of metaphorical assertions, a complex analysis of the way readers and listeners respond to these false assertions of identity. (...) It is not a matter of structure but of effect, and the study of metaphor should be a study of response.
>
> This might well be a fruitful line of inquiry. It would involve treating the notion of metaphor as a description of certain interpretative operations performed by readers when confronted by a textual incongruity, such as the assertion of a patently false identity. (...) the figurative is the name we give to effects of language that exceed, deform, or deviate from the code; codifications of previous excesses, deformations, and deviations only create opportunities for new turns. (Culler 1981: 208-209)

After Culler, it would seem in order to leave aside for the present the linguistic study of metaphor, and make it into a study of response. The electrocortical effects of textual incongruity, deformations and deviations would promptly highlight the predictions of the anomaly model that metaphor evokes a kind of shock effect. Recall that 'the poet must arouse a reaction and yet impede it, creating a tension in our nervous system sufficient and rightly calculated to make us completely aware that we are living something-and no matter what' (Eastman, *The Literary Mind*: 205, quoted in Richards 1965: 124).

The anomaly model is different from the comparison and interaction model in that metaphor processing supposedly involves an 'anomalous moment', materialized by a form of cognitive tension (the 'shock' at the 'clash of meanings', Beardsley 1982: 267). Thus, the shift from literal to figurative interpretation could be accompanied by the aforementioned ERP called N400, which is susceptible to anomalous sentence completions. Since in the Kutas & Hillyard studies, the N400-effect enhanced when the anomaly became more severe, literal expressions should evoke the smallest N400 amplitude, anomalies the highest and metaphors something in between.

If N400 marks the shift to the activation of figurative features, it may be a function of the distinctive literal features. Since the response criterion q_l is not met, too small a literal shared set leaves too many literal features distinctive. It could be inferred that a potential N400-effect is a function of too-large distinctive literal sets, remaining when the comparison phase (nearly) ends.

> So far almost nothing is known about the internal network of verbal communication and, in particular, about the neural stage in the output and input of distinctive features; let us hope that in the near future neurobiology will provide an answer to this question of primary interest for the comprehension and further study of the ultimate linguistic units. (Jakobson 1971: 688-689)

Twenty-five years have passed since Jakobson hoped for neuropsychological evidence on language processing. During this period, the neuropsychological research field has broadened to include normal readers, thereby refining its methods. For instance, Miall is quite optimistic when he reviews neuropsychological research to:

> ground (...) reader response issues on a firmer basis, on functions of the cognitive and emotional system about which reasonably clear neuropsychological evidence is now available. (...) the aim [is] to draw an outline of those aspects of current neuropsychological thinking that may illuminate the nature of literature and literary response. (...) Where reliable indicators of [foregrounding] exist, their neurological correlates can be assessed using EEG and EMG (electromyographic) measures. (Miall 1995)

Hopefully not overweeningly optimistic, an EEG-experiment was conducted in Section 7.4 on the N400-effect of literary metaphor processing. However, to appraise the import of the results, the indexing status of the N400-effect should be reviewed first, which is critically done in the next Section. One point that needs to be clarified before the N400-literature is discussed is the relation between cognitive brain potentials and arousal.

Berlyne's arousal theory in literature

In the theory of literature, Berlyne's (1960, 1971) arousal theory on artistic stimuli is often cited to illustrate the difference between literary and nonliterary reading on a physiological level.

> When a stimulus pattern fails to agree with an expectation that was aroused by what preceded it, we call it 'surprising'. (Berlyne 1971: 145)

The literary theorist Šklovskij (1965, originally 1917) influenced Berlyne's formulation of his surprise theory for artistic stimuli. Šklovskij argued that artistic texts are harder to process, owing to the higher frequencies of linguistic deviations (rhyme, semantics), which increase the complexity of the text or surprise the reader. Berlyne transformed these ideas into a theory of arousal states, in which the level of arousal would correlate with the hedonistic feelings of the subject.

> Any diminution of arousal due to temporary liberation from the restrictions of rationality can, we must suppose, counteract the arousal-raising potentialities of the anomalies that result. Consequently, the necessary conditions for the arousal boost or the arousal boost-jag can be realized. When we come across absurdities and incoherences in situations that are not clearly labelled artistic, playful, or humorous, we are apt to find them not enjoyable but profounding disquieting. The attendant arousal increment is too strong for an arousal boost, and there is no prompt relief to make possible an arousal jag. (Berlyne 1971: 171)

Berlyne's theory is an example of a psychological theory of cognitive-energetic processes. Studies on alertness in relation to the function of the recticular formation in the brain prompted the notion of a unidimensional arousal system, in which an optimum of arousal - not too little, not too much - would lead to optimal performance (see for an review of arousal studies Sanders, in press, ch. 10).

Berlyne envisioned the relation between level of arousal and aesthetic experience as an 'inverted-U' curve. One postulate underlying this view is that the relation between arousal and cognition is aspecific. In the unidimensional arousal theory, energetic and cognitve aspects of information processing are strictly separated (Sanders, in press, ch. 10). As a consequence, arousal may be evoked by all kinds of 'surprising' stimuli, without specific relevance for the processes involved. Among the various proposed measures of arousal are background EEG, muscle tension, galvanic skin response, heart rate, blood pressure, and pupil diameter (Sanders, in press, ch. 10). These measures - although sometimes yielding contradictory results - may reflect energetic aspects of stimulus processing. However, they are not informative with regard to the computational aspects of a process, as in the case of metaphor comprehension.

To conceive of the energetic effect of metaphor in terms of a 'shock' (Henle 1966: 182), then, might be a misconception. Instead, these effects may reflect computational procedures of metaphor processing. Metaphor is thought to be a simple mismatch of features, the effect of which is not an emotionally tinged arousal effect or 'shock'. Instead, the effect may depend on the outcome of computational procedures, so that arousal is not the proper measure.

This objection may also be advanced against other ideas on the effects of literary devices. It is common practice both in the theory of literature and in the psychology of aesthetics not to make a distinction between energetic and computational aspects of processing. Kreitler & Kreitler (1972), for instance, posited that an organism strives for an equilibrium (homeostatis). The reduction of psychophysiological tension is a satisfying experience, so that moderately deviating stimuli are optimal for anticipating tension relief, and are supposedly much appreciated in the perception of art and literature. However, despite the ample speculations on arousal effects for aesthetic stimuli, there is no experimental evidence that in literature, arousal is experienced as 'more pleasant' than in trivial prose.

Presumably, most of the aesthetic pleasure in literature is given by stylistic deviation. Berlyne called these the 'short-term novelties', which concern phenomena such as metaphor, synaesthesia, polyvalence (ambiguity), rhyme, and orthography. However, the deviateness of these stimuli is paramountly based on computational processes, not on different levels of energetics. Muscles do not contract when encountering a metaphor. The heart does not pound in the throat because of synaesthesia. Blood vessels do not widen when confronted with a lexical ambiguity. One does not get sweaty due to a surprising rhyme, so that arousal states are noninformative on the functioning of these literary devices.

Therefore, the electrocortical effect of metaphor processing is investigated with the cognitive ERP called N400. In the next Section, a review is offered on the sensitivity of the N400 to semantic stimuli. Moreover, the limitations of ERPs for language research are pointed out.

7.2 **Psychophysiology: Studies on semantic deviations**

Since the original EEG research of Berger (1929) and Adrian & Matthews (1934), the idea has been elaborated that electrical brain activity corresponds with the processing of internal and external events. The so called 'evoked response potentials' (ERPs) are supposed to reflect brain activity brought about by a stimulus. The ERP is thought to provide a continuous record of information processing in the brain, whereas behavioral measures (e.g., RT) are the end product of such processing. It can be deduced from a combination of RT and ERP that systematic changes take place in both processing duration and brain activity as a result of systematic manipulation of variables. The precise nature of the electrocortical activity is unknown. However, the observed effects can be used to formulate hypotheses about the processes involved.

On the other hand, the predicted modification of an ERP is not necessarily a direct manifestation of the presumed process. There is no way of telling whether an ERP reflects one or more processes, or which process it reflects. It is also unclear which part of the morphology (onset, flank or peak) of an ERP waveform is most relevant to the processes under study. Thus, results are correlative rather than causal, and should be treated as description rather than explanation (cf. Meyer, Osman, Irwin & Yantis 1988).

Despite these complications, recording EEG at different scalp locations may help in suggesting different functional areas of the brain involved in the variations of the stimulus (cf. Nunez 1981). Therefore, ERPs may form a complement to gain insight into the organization of mental processes. The analysis of ERPs has been useful in the study of stimulus detection and recognition, memory, reaction processes, attention and language (Hillyard & Kutas 1983).

With respect to language, Kutas & Hillyard (1980_a , 1980_b , 1981, 1983) dealt with ERP components in response to semantic anomalies. Words that were semantically incongruent with the sentence context ('he spread the warm bread with socks') evoked negative electrical activity (N) around 400 ms after onset at centro-parietal locations. The N400-amplitude rose in response to words that were rated as highly anomalous, whereas moderately anomalous words elicited a lower amplitude. Kutas & Hillyard stated that N400 is probably susceptible to the relation of a word with the priming by its context. Kutas, Lindamood & Hillyard (1984) claimed that amplitude and latency of the N400 varied as a function of factors such as presentation time, reading competence and sentence context. Boddy & Weinberg (1981), Bentin, McCarthy & Wood (1985) and Rugg (1985) confirmed that the height of the N400-amplitude was sensitive to the mismatch of semantic expectations, although phonological mismatches were affecting the amplitude as well.

Fischler, Bloom, Childers, Roucos & Perry (1983) tested appropriate and inappropriate instance-category comparisons that were stated affirmatively or negatively: 'A sparrow is a bird', 'a sparrow is a vehicle', 'a sparrow is not a bird', 'a sparrow is not a vehicle'. Subjects responded 'true' or 'false', while RT and EEG were recorded. When categories mismatched, a large negative going wave was found between 300-500 ms. This N400 occurred irrespective of the truth value of the proposition (cf. 'a sparrow is not a vehicle'). False propositions about matching semantic categories ('a sparrow is not a bird') evoked less N400-activity. According to Fischler et al., comparison of semantic categories occurred prior to a conclusion about the truth value. This pattern was also found in the RTs. False propositions with matching categories were processed faster than true propositions with mismatching categories.

Chwilla, Brown & Hagoort (1995) investigated whether N400 is a reflection of automatic or controlled processing. In a semantic priming paradigm, word pairs were presented, the second of which was a real word or a nonword, written in lower or upper case. Moreover, word pairs could be related or not. Two tasks were performed: A lexical decision task (word/nonword?) and a typographical discrimination task (lower/uppercase?). It was expected that a lexical decision task would involve automatic lexical access (is it a word?) as well as controlled lexical integration (do words match?), whereas a typographical task would only involve automatic lexical access. Therefore, if N400 is a reflection of controlled processes, it should be present in the lexical decision task, and not in the typographical discrimination task. To vary the depth of semantic processing, two stimulus lists were exploited, one with a high proportion (80%) of related word pairs, and one with a low proportion (20%). The high-proportion list was thought to encourage deeper semantic processing of the first word, because this would lead more often to successful predictions of the second word than in the low-proportion list. Accordingly, N400 should be higher in the high-proportion list than in the low-proportion list.

Three subject groups joined an RT/ERP experiment. The first group performed a lexical decision task (is it a word/nonword?), and the second a typographical discrimination task (lower/upper case?). The third group performed a passive reading task. This was done to control for unwanted side-effects of performing a decision task, which are not present in a natural reading situation.

In the typographical task, no RT differences were found between words and nonwords, nor between related versus unrelated word pairs. In the lexical decision task, RTs to words were faster than to nonwords, and related word pairs were faster than unrelated, especially in the high-proportion list. N400 occurred in the lexical decision task as well as in the silent reading task, whereas it remained absent in the typographical task. Larger amplitudes were observed for unrelated than for related word pairs. The high-proportion list yielded higher N400-amplitudes than the low-proportion list.

It was concluded that N400 reflects controlled lexical integration rather than automatic lexical access, because in the typographical task - which assumingly only required lexical access - the N400 was not evoked. Moreover, N400 may be sensitive to deeper forms of processing, because in the lists with high proportions of related word pairs, the N400 to unrelated word pairs was higher than in the low-proportion list.

Thus far, it seems that the N400 was connected to semantic mismatches in a relatively unproblematic way. It was insensitive to typographical entities of the words (lower/upper case), to logical constraints (affirmative or negative sentences), or to the position of the aberrant word in the sentence. In addition, N400 seemed to be susceptible to semantic relatedness of words on a deeper level of processing (lexical integration).

However, other evidence suggested that N400 is also evoked by non-linguistic stimuli (pictures), and sometimes by word properties that are not necessarily semantic (phonology, orthography). Rugg (1984) designed a rhyme-matching task in which rhymes attenuated the N400-amplitude, whereas nonrhymes increased it. Additional evidence was found by Hoorn (1996). Rugg & Nagy (1987) found that nonwords with incorrect orthography for the standard language evoked smaller N400-amplitudes than orthographically correct nonwords. Stuss, Picton & Cherri (1986) identified N400-activity for naming pictures, Nigram, Hoffman & Simons (1992) for sentence endings with unexpected pictures, and Stuss, Sarazin, Leech & Picton (1983) for mental rotation of geometrical figures. Although it may be argued that these are also semantic stimuli, these findings call for careful consideration of the linguistic status of the N400.

7.3 A moderate electrocortical effect for metaphors? Right hemispheric?

Section 7.1 ended with the observation that electrocortical effects of metaphor processing are presumably related to computational procedures. Therefore, the 'shock' effect of metaphor should not be seen in the light of a unidimensional arousal theory of surprise, but rather as a modulation of the N400-amplitude, which is measured in response to semantic mismatches (Section 7.2).

If the N400-effect is the manifestation of semantic mismatch on a controlled level of cognitive processing, an increase of the amplitude should be found when words are less related, as in the case of metaphors. Thus, literal expressions should show less N400-activity than metaphors, whereas anomalies should show most.

Moreover, there may even be a hemispheric specialization for processing the diverse expression types. Danesi (1989) was among the first to hypothesize about the locus of metaphor processing in the neurological anatomy. He speculated that different stages in metaphor processing are connected with different hemispheric specializations. Danesi claimed that processing the *B-term* of a metaphor is a right-hemispheric activity, whereas the *A-term* is processed in the left hemisphere. These views were based mainly on the results of two studies: Goldberg & Costa (1981) and Winner & Gardner (1977).

Goldberg & Costa suggested that the anatomy of the right hemisphere is better fit for processing novel stimuli, because the neural-synaptic connections supposedly are more numerous and intertwined. The organization of the left hemisphere was viewed by these authors as a more sequential neural-synaptic structure, which is better equipped for processing overlearned patterns.

Inspired by the literary theorist Jakobson (1960), Winner & Gardner questioned whether the left hemisphere was specialized for figurative language. They presented metaphors to a group of subjects with left-hemisphere damage, a group with right-hemisphere damage, and a normal control group. For paraphrasing the metaphors, subjects could choose from four pictorial interpretations. It was concluded that normal controls and left-hemisphere damaged patients could discriminate better between literal and metaphoric interpretations than right-hemisphere damaged patients. As Danesi put it, this was 'for the first time in neuropsychology [that] an experimental link was established between the right hemisphere and the content-structure of a metaphor.'[1]

Based on these studies, Danesi suggested that metaphor may be viewed among the class of novel stimuli, given the fact that the *B-term* is an unexpected completion of the *A-term*. In other words, the *A-term* could be conceived as an automatically activated pattern, which is encoded in the left hemisphere; the *B-term*, on the other hand, as a novel stimulus, which is encoded in the right hemisphere. The *C-term*, then, is the synthesis of *A-* and *B-term*, in which probably both left and right hemisphere are involved.

If a metaphoric *B-term* is a novel stimulus in combination with the *A-term*, it follows that there should be asymmetrically distributed N400-activity (cf. Kutas & Hillyard 1980_b , 1982). If novel stimuli are predominantly processed in the right hemisphere, the N400-activity for metaphors and anomalies should be foremost at right scalp locations, whereas for literals, N400 should be biased towards left scalp locations.

Kutas & Hillyard (1980_a) found that the N400 was larger for strong than for weak anomalies; hence literal expressions should evoke small effects, metaphors medium effects and anomalies large N400-effects. In line with Danesi, the effects should be left-hemispheric for literals and right-hemispheric for metaphors and anomalies.

7.4 ERP experiment on the N400-effect of literary metaphor

Subjects were female undergraduates of language and literature (N = 24), Dutch native speakers, aged between 19 and 27 years old. They received pay for volunteering in a three-choice classification decision task as described in Section 6.4, Chapter 6. To avoid the repetition effects observed in the RT experiments of Chapter 6, the 3-choice task - with the least repetitions of stimuli - was preferred to the 2-choice task.[2] However, Chapter 6 indicated that results depend highly on the design, so that the outcomes of this experiment should be valuated in this light.

Subjects were tested during the second week following the first day of their monthly period. They had never participated in an experiment, and received no information about stimuli or research aims. They were right-handed and had normal or corrected to normal vision.

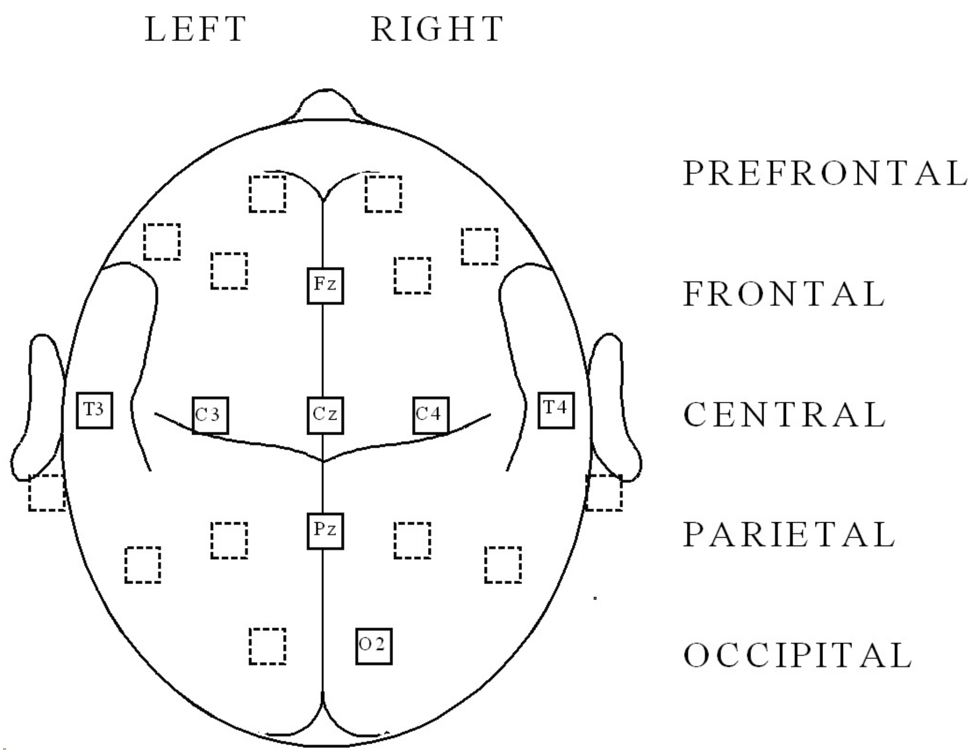
Stimuli were the expressions of Table 4.1, Chapter 4. Trials were built up, according to the RT experiment in the previous Chapter. They consisted of a sequence of prime - fixation line - target. In the *expression* condition, the prime consisted of an expression without *B- term* (the child is a), after which a fixation line was projected (the child is a ____). The width of the fixation line had four or five underscores, corresponding to the number of *B-term* letters. Targets were *B-terms*, completing the expression (the child is a moon). Unlike the RT experiment of Chapter 6, reading the prime was not self paced, because the software package (InstEP, Campbell & Bell 1992) furnished no options to interact with the subject. In the *expression* condition, the prime was presented two seconds, followed by a one-second fixation line, which was replaced by the *B-term* for two more seconds. The inter-trial-interval was also two seconds.

The duration of the primes in the *context* condition was based on a mean reading time of two words a second. The build up of the trials in *context* was different from the RT experiment in Chapter 6. Since the readers could not indicate when they finished reading the text, it might be that the *B-term* was projected while being in the middle of a sentence. Therefore, the context without the *B-term* was presented during the mean reading period, whereafter the screen was cleared for one second, and the context returned. This indicated that they should read the expression that preceded the *B-term*, for which they had two seconds. Then, the fixation line (1 second) and *B-term* (2 seconds) were projected in the context. Warning feedback was provided acoustically by the session leader.

Before starting the experiment, subjects received 10 practice trials on 3 prototypical instances of each expression type, so as to define literal expressions, metaphors, and anomalies. Subsequently, subjects had 20 practice trials of unrepeated expressions per expression type in both conditions, until the standard deviation of the reaction times was less than 15% of the mean, and errors were less than 10% of the total number of responses (Sanders 1980). The practice stimuli were unrelated to those of the actual test set.

Regarding apparatus and recordings, the software package InstEP (Campbell & Bell 1992) ran on a 368SX slave and 486DX2 master computer, controlling stimulus presentation and data acquisition (EEG, VEOG, HEOG, movement time (MT), RT and decisions). Subjects were lying on a bed in a sound-attenuated dimly lit (15 lux) cubicle. A response-button box with three decision keys and a home key was positioned at an optimal distance for each subject, half the trial block at the left side, half at the right side of the subject. Each decision key was assigned to one of the options: L (literal), M (metaphor) or A (anomaly). Stimuli were presented on a screen placed at a distance of 75 cm from the subject. HEOG and VEOG were measured by two tin electrodes placed at the outer canthus of each eye and by two tin electrodes placed infra-orbital and supra-orbital in line with the pupil. EEG was recorded from nonpolarizable tin electrodes mounted in an elastic cap (Electro Cap International) and located at standard left and right hemisphere positions, spanning the surface of the scalp (Figure 7.0: International 10/20 System names (Jasper 1958): Fz, Cz, C3, C4, Pz, T3, T4, O2). Linked ear electrodes were used as reference electrodes.

Figure 7.0: Electrode locations at the scalp, according to the International 10/20 System.



EEG was sampled continuously, because InstEP could not sample single trials that incorporated more than two stimuli. The signals were sampled with a lowpass filter of 30 Hz and a time constant of five seconds, and were digitized at 100 Hz. Electrode impedance never exceeded 3 kΩ.

Since InstEP could not sample two response times (MT and RT) for one trial, MT was recorded directly from the decision keys, while releases from the home key were registered via an extra EEG-channel. The latency from *B-term* onset to the polarity shift at home key release was considered to reflect RT.

The design consisted of three between-subject factors of condition order (*expression* before *context* vs *context* before *expression*), decision key order (LMA, LAM, MLA, MAL, ALM, AML) and index finger order (right before left vs left before right) with two within-subject factors of condition (*expression* vs *context*) and expression type (literals vs metaphors vs anomalies). Condition was blocked and all expressions of Table 4.1 were pseudo-randomly mixed within blocks.

In the same way as in Chapter 6, Section 6.4, RT did not represent completed decisions. Subjects probably used all time available to make their decisions. Therefore, all further analysis was performed on MT+RT, or arrival time (RT_a) for short.

Results for arrival times

Analysis of the 2*6*2*2*3 design would result in N = 1 for each cell, so that within-cell variation is absent. Therefore, it was first investigated whether certain factors could be excluded from the analysis. The previous Chapter showed that condition order affected RT_a , so that two preliminary MANOVAs were carried out, one excluding decision key order, and one excluding index finger order.

MANOVA without decision key order showed that effects of index finger order were insignificant (e.g., index finger order by condition order by condition by expression type: Pillai's Trace = .07, $F_{2,19} = .80$, p = .462; main effect of index finger order: $F_{1,20} = .43$, p = .521). MANOVA excluding index finger order showed that effects of decision key order were insignificant (e.g., decision key order by condition order by condition by expression type: Pillai's Trace = .52, $F_{10,24} = .86$, p = .576; main effect of decision key order: $F_{5,12} = 1.34$, p = .313). Since factors were counterbalanced, order of decision key and index finger were pooled over subjects, so that the design was reduced to a 2*2*3 of condition order, condition and expression type.

Figure 7.1: Grand mean arrival times (ms) in a 3-choice task for literal expressions (■), metaphors (▲) and anomalies (●) in *expression* and *context* (correct responses). Solid lines signify *expression* before *context*, dashed lines *context* before *expression*. In each cell, N = 12.

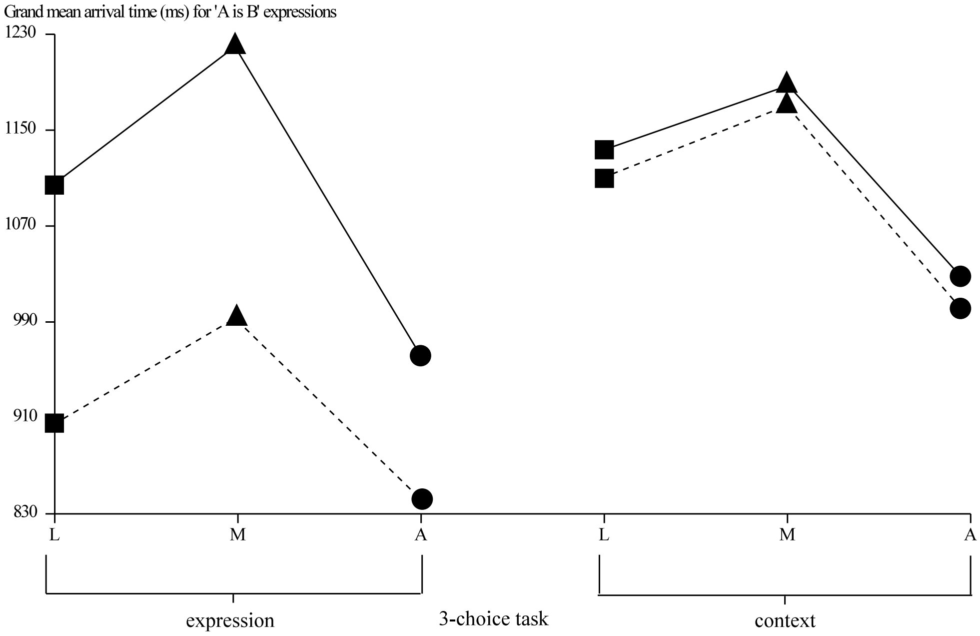


Figure 7.1 displays the grand means for the arrival times. The appropriate statistics are found in Table 7.0 (Appendix), and underscored the next statements:

(I) Performing *expression* as the first condition (solid) resulted in slow responses, while performing *expression* last resulted in fast responses (dashed). In *context*, the order effects faded (interaction of condition order by condition). (II) *Expression* was performed faster than *context* (main effect of condition). (III) Anomalies were reacted to fastest - followed by literals - whereas reactions to metaphors were slowest (main effect of expression type).

Identification of ERPs

For EEG-analysis, InstEP files were converted to the STPBS format (Woestenburg 1994). Horizontal and vertical eye movement artifacts were removed by a regression analysis in the frequency domain.[3] The presence of eye movement was tested in a covariance analysis and significant parts were deleted. Clipped trials were detected and replaced by an interpolation between two adjacent trials of the same category. The lower limit of outliers was set by the mean plus two times *sd*. Trials exceeding this lower limit were interpolated. Subsequently, the Orthogonal Polynomial Trend Analysis

(OPTA: Woestenburg, Verbaten, Van Hees & Slangen 1983; Woestenburg 1994) was performed for each subject. A complex variance analysis in the frequency domain tested which frequencies were obscured by noise and which frequencies contained strong ERPs. Upon these statistics, it was decided which insignificant frequency bands were filtered. Although the OPTA permits trend analysis, no trend was expected, so that only the mean was used for filtering.

Time slice analysis

Next, statistical analysis was performed over all subjects by dividing the ERPs elicited by the *B-term* into subsequent time slices (Woestenburg, Das-Smaal, Brand & Kramer 1992). A sample period of 20 ms was used for a window extending 0-800 ms after *B-term* onset with a 150 ms baseline. A priori, it was expected that N400 occurred in the time span between 265-445 ms after *B-term* onset. The 10 time slices of 20 ms each in this area were tested for significant amplitude differences, all having to meet an α-level of .05. To arrive at a more conservative rejection area, .05 was divided by 10 (the number of time slices). Hence, * = $p < .05$, ** = $p < .005$. The time slice analysis was used as an explorative analysis to identify the amplitude differences at each time interval for the effects of condition and expression type only. This led to selecting the amplitudes of two time intervals (concerning the P300 and the N400), on which data a MANOVA was run for the complete design.

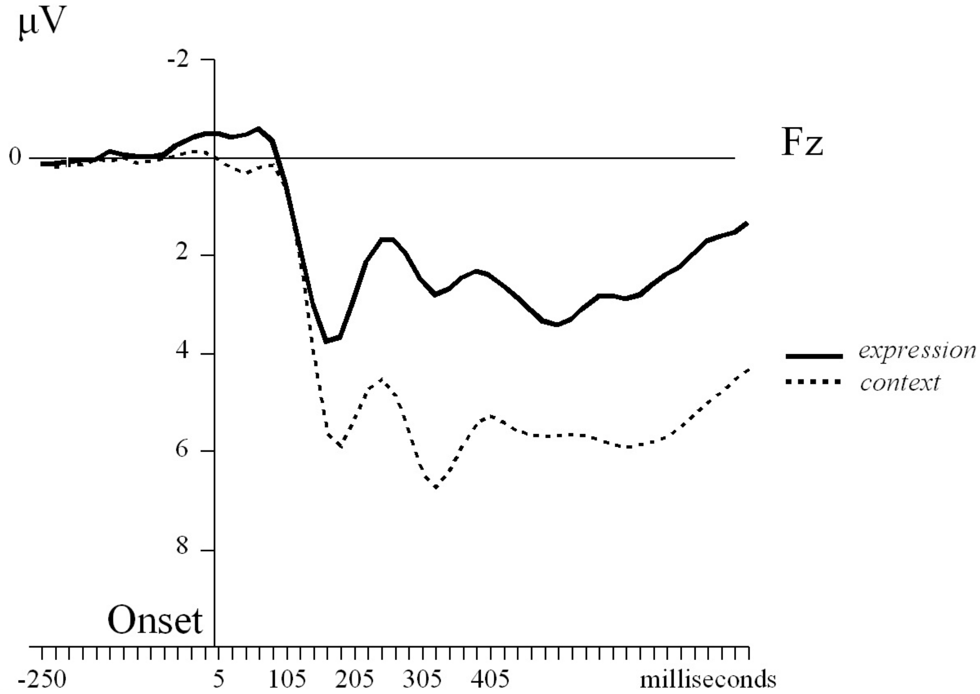
Results of the time slice analysis

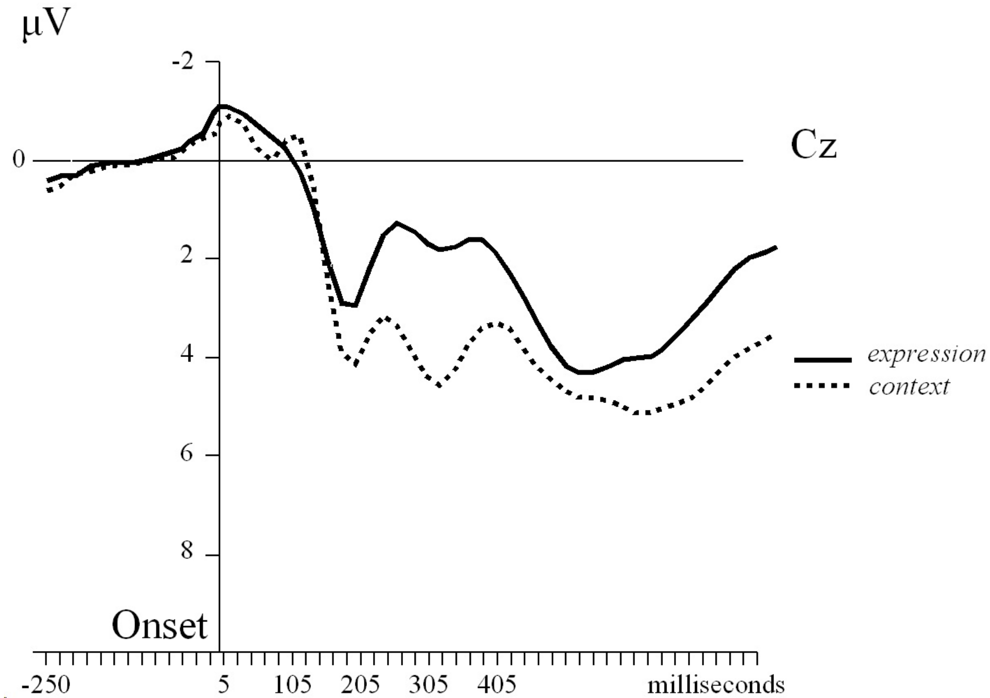
Figure 7.2 up to 7.3 show the main effects of condition and expression type and their interaction, while Table 7.1 (Appendix) shows the maximal amplitude differences between 265 and 445 ms after *B-term* onset.[4] Notice that the maximal differences occurred at the rising flank of the N400, rather than at a fixed moment in time. This raises the problem that it is uncertain whether or not the same ERP was measured for each factor or factor level. This issue will be elaborated in the MANOVA analysis.

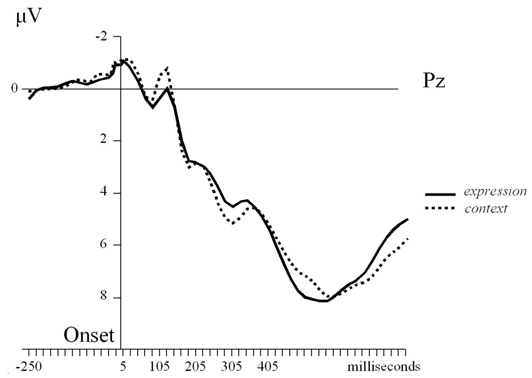
Taken together the results of Figures 7.1 up to 7.3 and Table 7.1 (Appendix), the following effects emerged 265-445 ms after *B-term* onset:

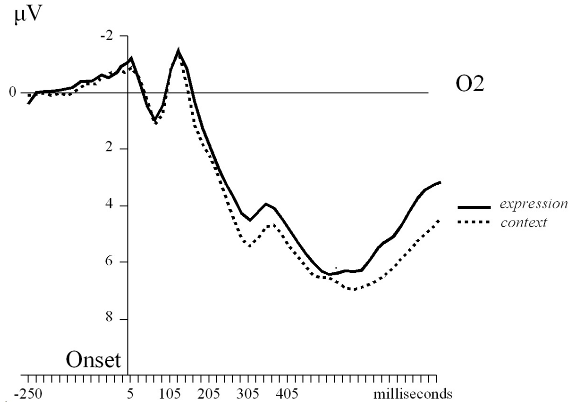
(I) *Expression* evoked larger negative deflections than *context* at Fz, Cz, C3, C4 (Figure 7.2, main effect of condition). (II) Anomalies and metaphors elicited larger negative shifts than literals at all electrode locations. Particularly at Fz, anomalies raised stronger amplitudes than metaphors, although this trend remained insignificant (Figure 7.3, main effect of expression type). (III) At Fz, anomalies in *expression* yielded larger negative shifts than metaphors, which yielded much larger amplitudes than literals. *Context* merely mitigated these differences. At the other locations, however, metaphors in *expression* invoked larger negative shifts than anomalies, which showed larger negative shifts than literals. *Context* not only had a mitigating effect, it also caused anomalies to show stronger negative deflections than metaphors (Figure 7.4). However, these effects were only significant at a liberal α-level, particularly for comparisons containing literals and metaphors at Fz, Pz, C3, C4, literals and anomalies at Pz, and metaphors and anomalies at OZ and C4 (Table 7.1: interaction of condition by expression type).

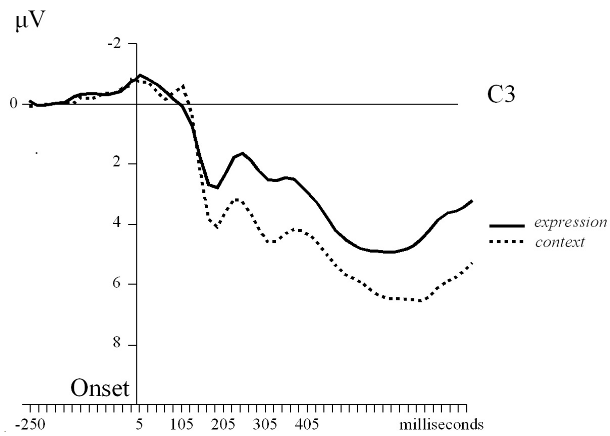
Figure 7.2: Main effect of condition on the grand mean EEG. Solid lines exemplify *expression*, dashed lines *context*. Negativity up. See also next page.






µV
-2
0
2
4
6
8
Pz
expression
context
Onset
-250
5
105
205
305
405
milliseconds


µV
-2
0
2
4
6
8
O2
expression
context
Onset
-250
5
105
205
305
405
milliseconds


µV
-2
0
2
4
6
8
C3
expression
context
Onset
-250
5
105
205
305
405
milliseconds

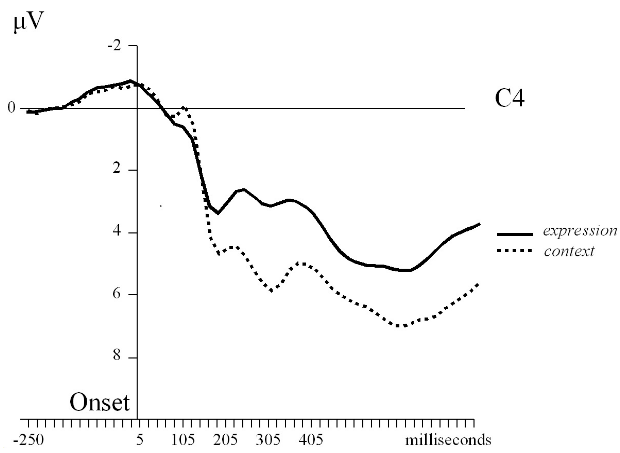
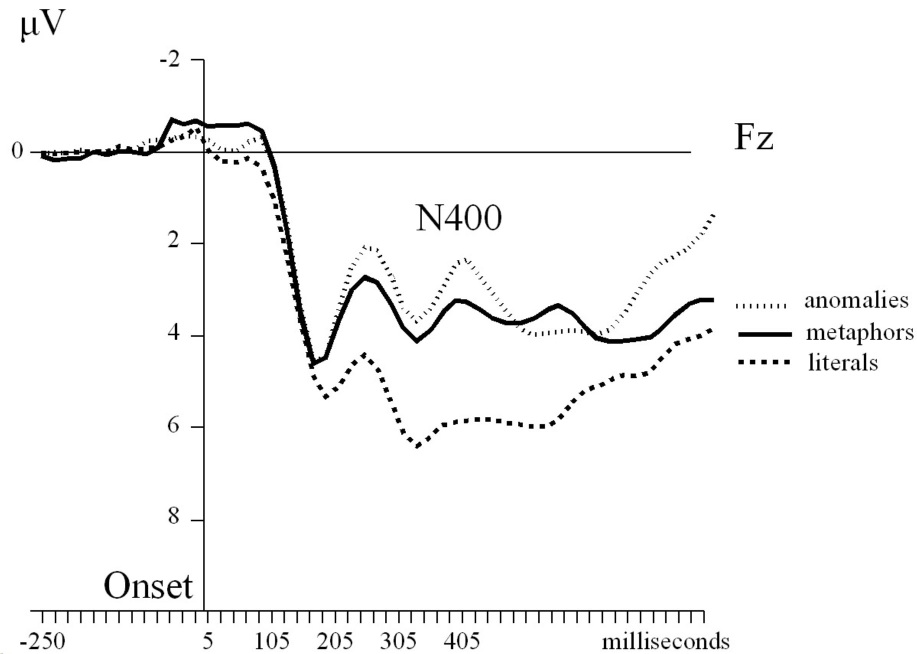
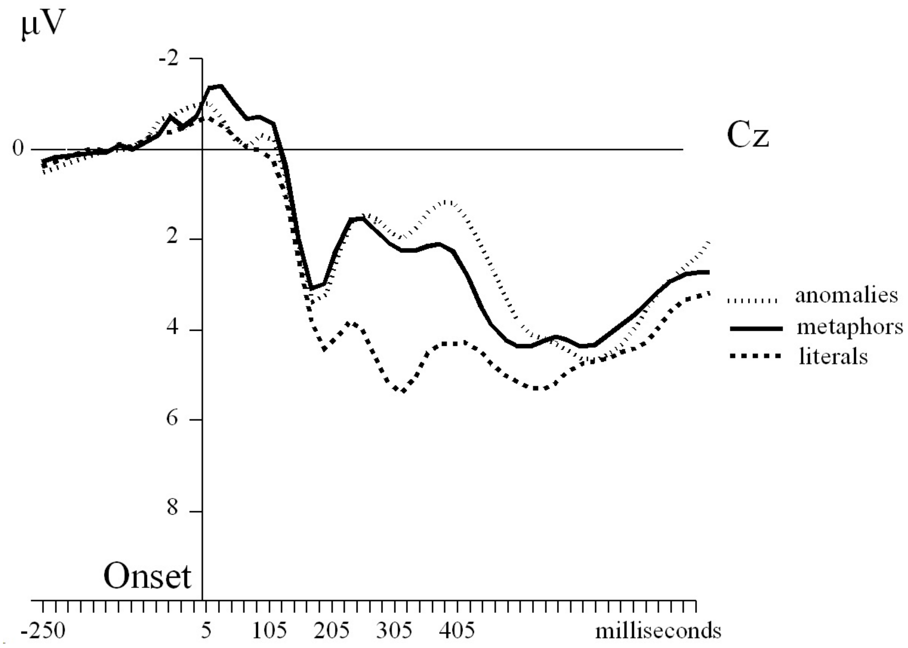
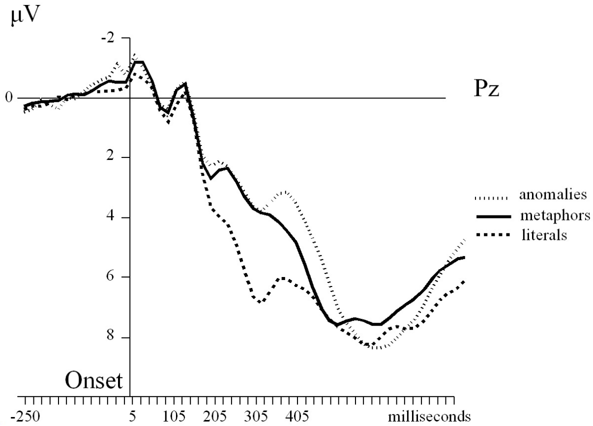
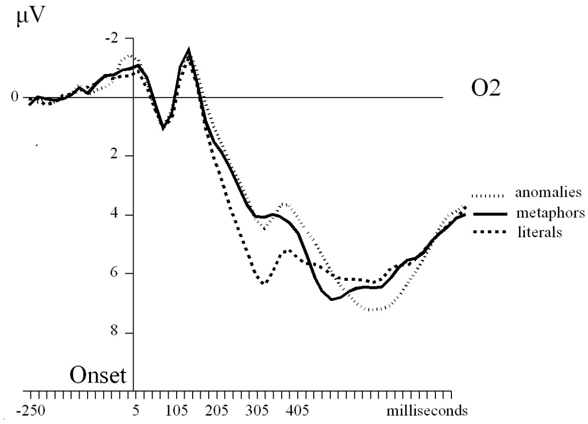

µV
-2
0
2
4
6
8
C4
expression
context
Onset
-250
5
105
205
305
405
milliseconds

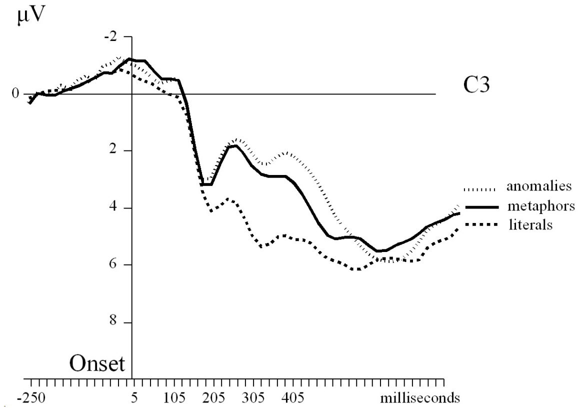
Figure 7.3: Main effect of expression type on the grand mean EEG. Literals are dashed, metaphors solid, and anomalies dotted. Negativity up. See also next page.






µV
-2
0
2
4
6
8
Pz
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


µV
-2
0
2
4
6
8
O2
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


µV
-2
0
2
4
6
8
C3
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

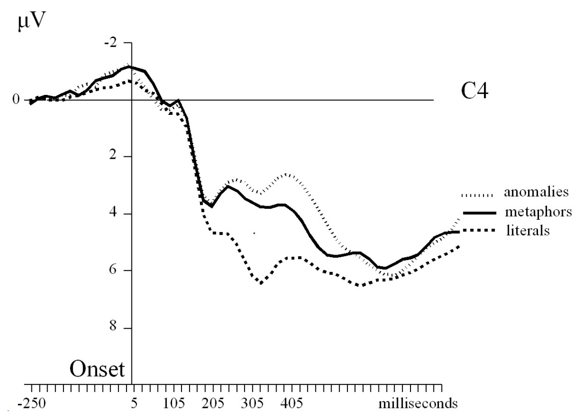
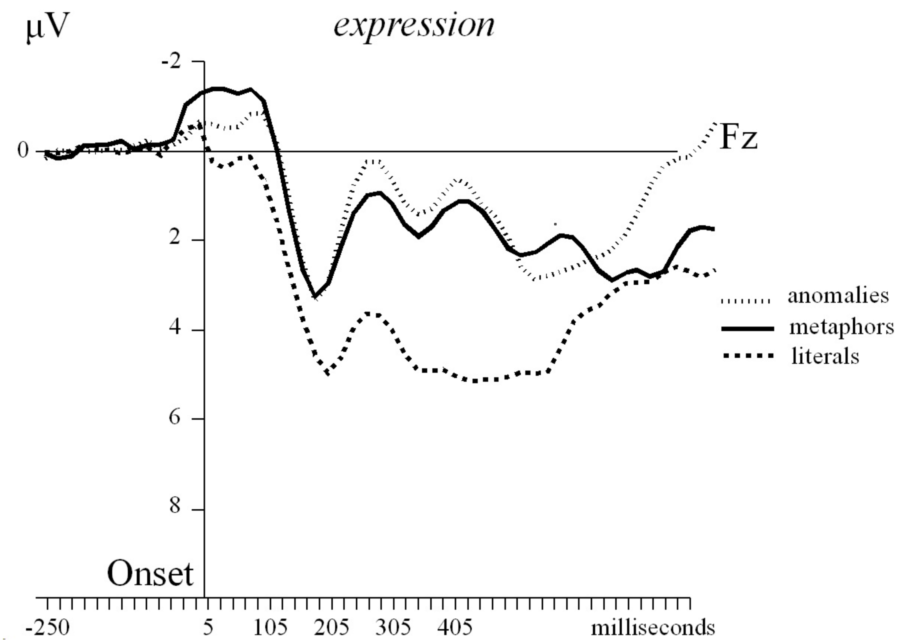
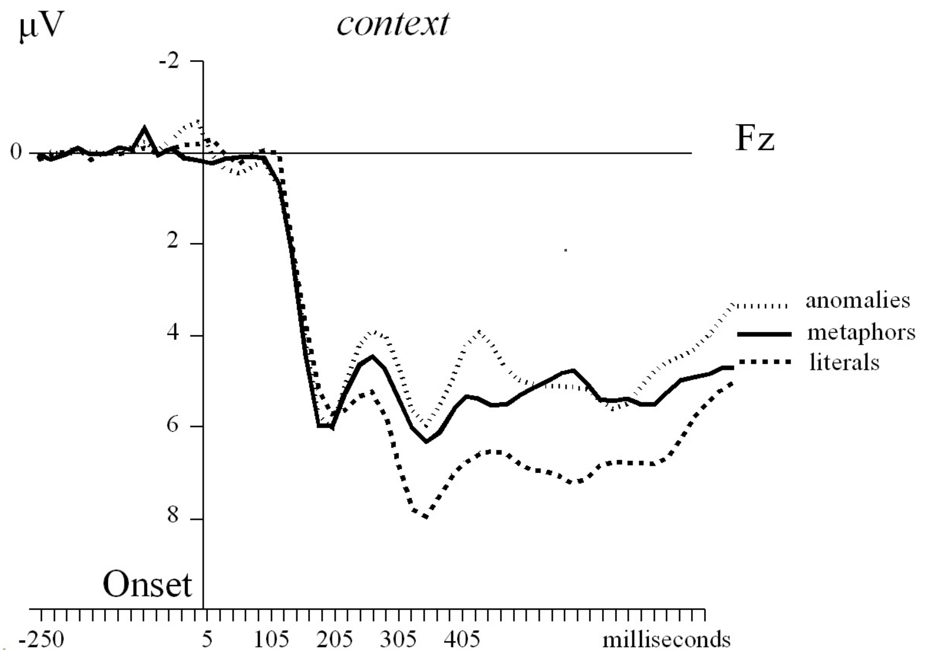
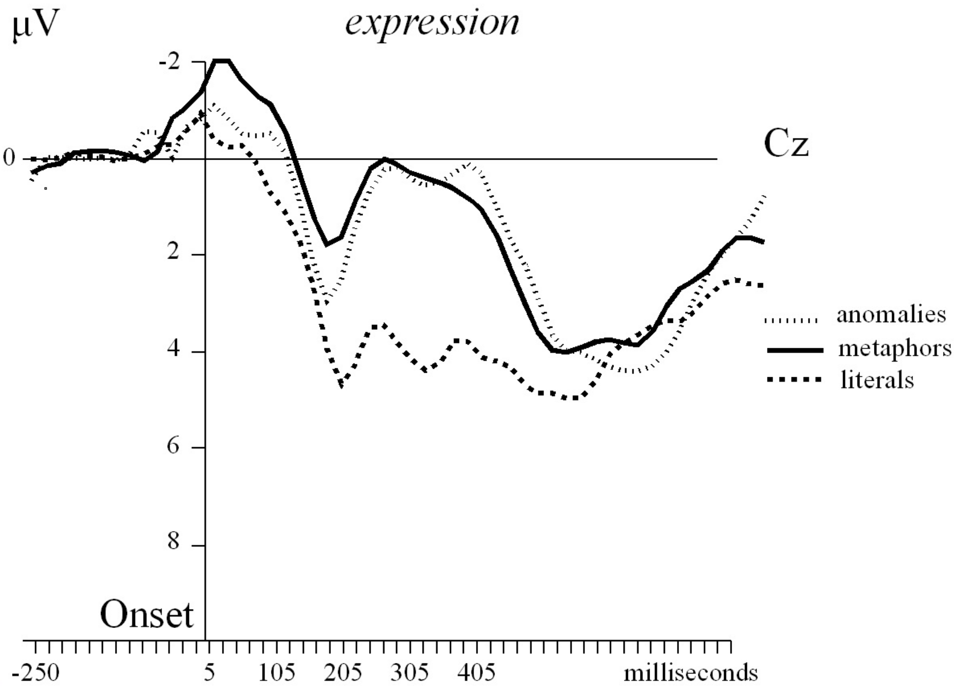
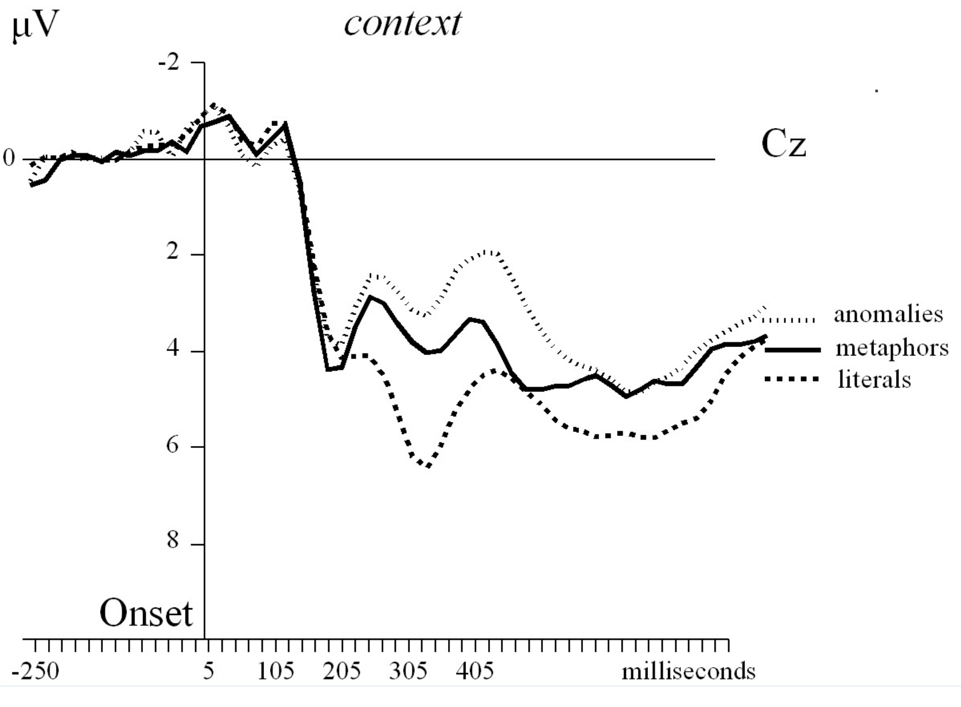

µV
-2
0
2
4
6
8
C4
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

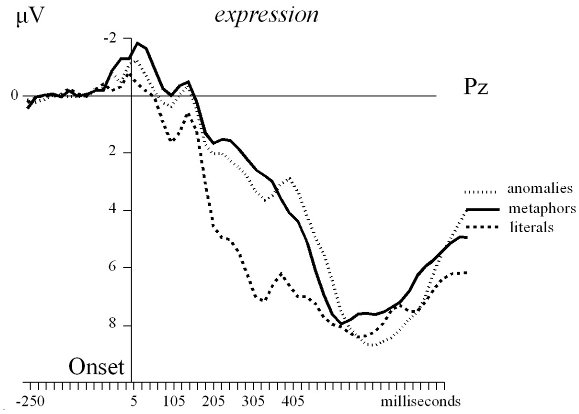
Figure 7.4: Interaction of condition by expression type on the grand mean EEG. Literals are dashed, metaphors solid, anomalies dotted. Two panels are shown per electrode location. Upper panel: Results for *expression*. Lower panel: Results for *context*. Negativity up. See also next pages.

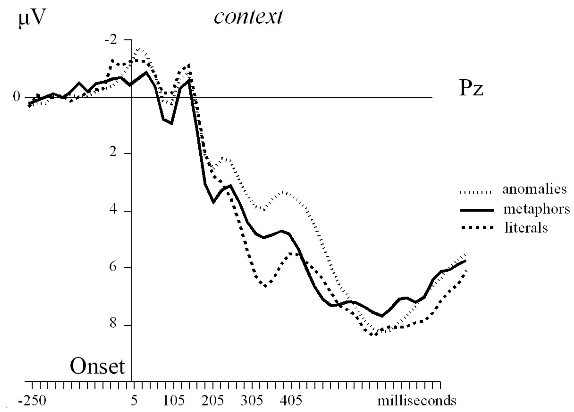


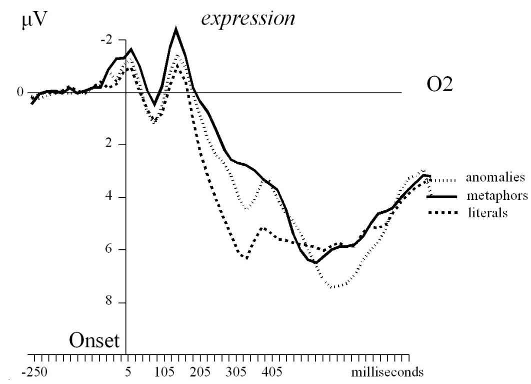


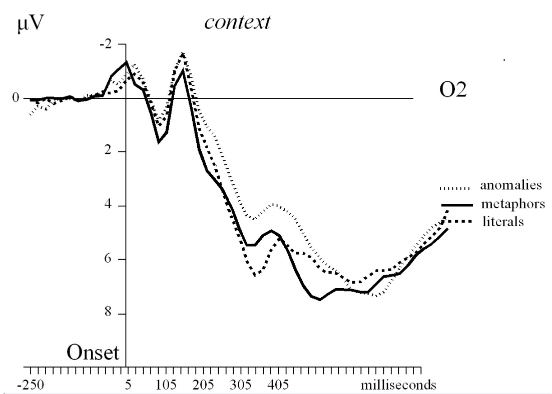

μV
expression
-2
0
2
4
6
8
Cz
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

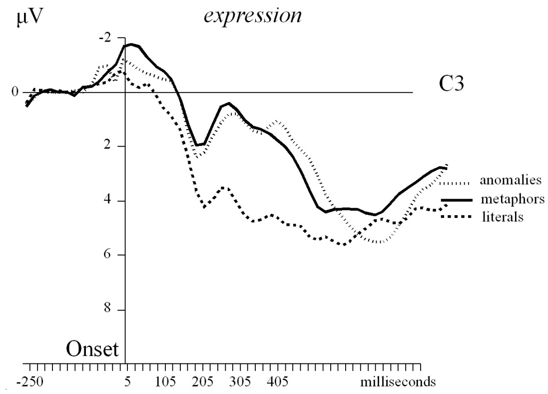

μV
context
-2
0
2
4
6
8
Cz
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

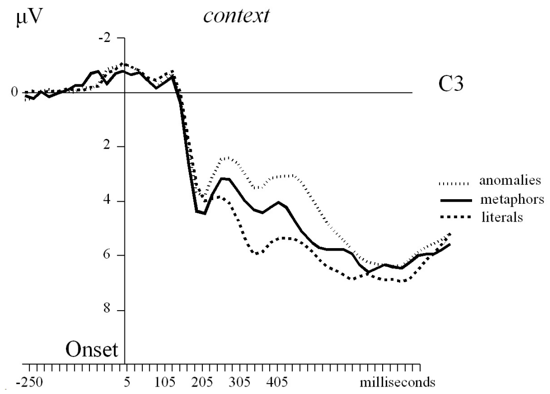

μV
expression
-2
0
2
4
6
8
Pz
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

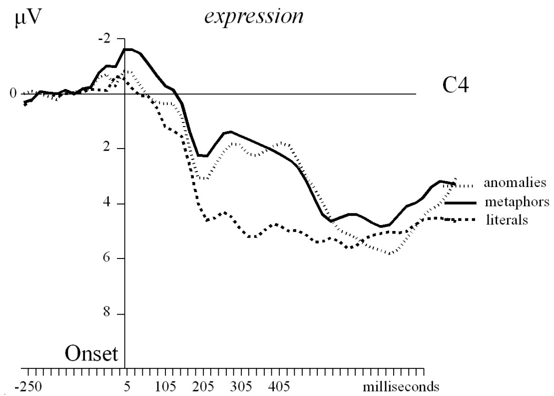

μV
context
-2
0
2
4
6
8
Pz
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

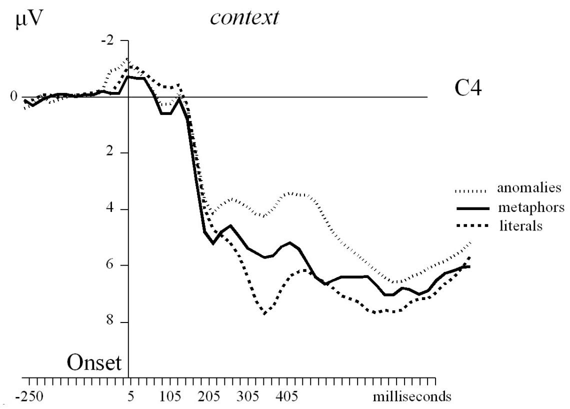

μV
expression
-2
0
2
4
6
8
O2
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


μV
context
-2
0
2
4
6
8
O2
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


μV
expression
-2
0
2
4
6
8
C3
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


μV
context
-2
0
2
4
6
8
C3
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


μV
expression
-2
0
2
4
6
8
C4
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds


μV
context
-2
0
2
4
6
8
C4
anomalies
metaphors
literals
Onset
-250
5
105
205
305
405
milliseconds

To look into the topography of the interaction in more detail, the amplitude of the grand mean ERP at 385 ms after *B-term* onset was scored at the midline as shown in Figure 7.5.

Figure 7.5: Grand mean amplitude, 385 ms after *B-term* onset at midline locations. Negativity up.

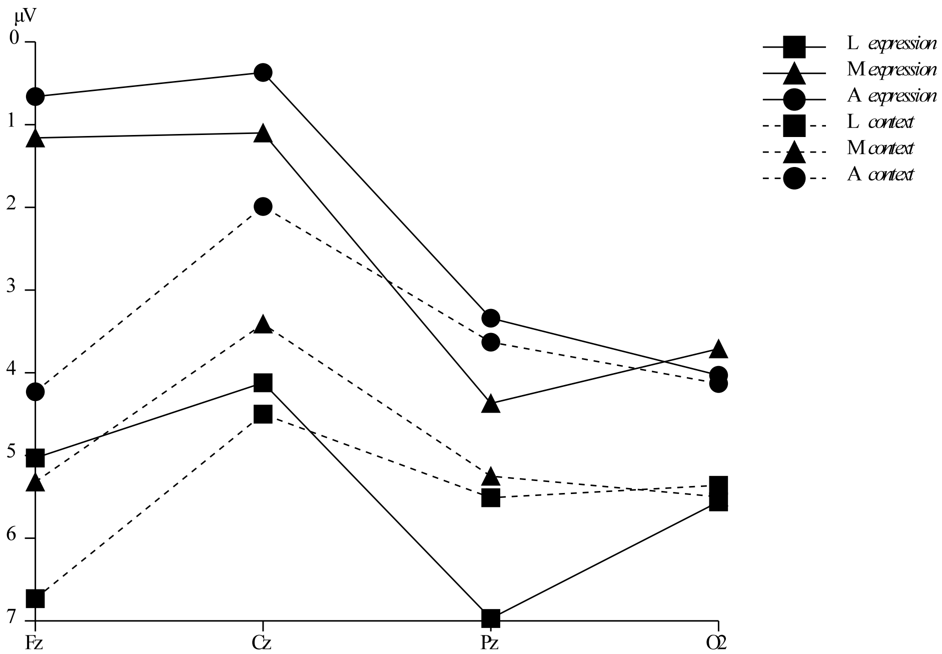


Figure 7.5 shows that (I) in *expression*, the smallest negative shifts were found at posterior locations (Pz, O2), whereas the largest were found at anterior locations (Fz, Cz). (II) In *context*, the smallest negative shifts were found anteriorly (Fz), and the largest posteriorly (cf. condition by lead, Appendix Table 7.3). (III) For both conditions, expressions reached their maximal amplitude at Cz.

MANOVA on P300 and N400

The time slice analysis covered a broad range of time intervals, in which different ERPs may be identified. Maximal amplitude differences were found for the rising flank of the N400, which were overlapping the P300 area. In other words, P300 also could be sensitive to processing the *B-term*.

Physiological data are repeated measures. If a factor has more than two degrees of freedom, MANOVA should be carried out (O'Brien & Kaiser 1985; Vasey & Thayer 1987). Since the N400 was evoked by linguistic stimuli, F-values with stimuli as the random factor should also be calculated (Clark 1973). Otherwise, generalizations are limited to the presently used stimulus set. However, the ERP method has the disadvantage that an ERP cannot be estimated for single stimuli, due to the unfavorable signal-noise ratio. Thus, the ERP is always a (grand) mean value that needs at least 20 observations. F-values with stimuli as the random factor, therefore, have to be cancelled, so that generalizations are limited to the stimulus set in question. In fact, the complete N400-literature suffers from this restriction.

For the same reason, it is impossible to analyze the effects of, for example, lexical ambiguity. Since the estimation of an ERP needs a minimum number of observations, even more inconveniences arise. Amplitude differences for the complete window (0-800 ms) may be computed for the main effects and first-order interactions. However, for implicit factors (decision key order, condition order, index finger order) too few observations remain to establish a stable ERP. Preferably, then, one would select the maximal amplitude differences to investigate the higher-order interactions in a MANOVA. However, this maximal amplitude difference should be fixed at one time interval for all factor levels, otherwise it is impossible to tell whether the same ERP applies. Unfortunately, in the present data, the maximal amplitude differences occurred at different moments of processing the *B-terms*.

It was decided, therefore, to carry out the analysis on the time intervals 305 ms (P300) and 385 ms (N400) after *B-term* onset, rather than on maximal amplitude differences. MANOVA was run on the 6*2*2*6*2*3*2 design of the between-subject factors: Decision key order (LMA vs ALM vs MAL vs MLA vs AML vs LAM), condition order (*expression* before *context* vs *context* before *expression*), index finger order (left before right vs right before left) and the within-subject factors of electrode lead (Fz vs Cz vs Pz vs O2 vs C3 vs C4), condition (*expression* vs *context*), expression type (literal vs metaphor vs anomaly) and time slice (305 ms vs 385 ms). The analysis of effects of decision key order by condition order by index finger order was prohibited by too few subjects. Therefore, three separate analyses were performed for the effects of decision key order by condition order, decision key order by index finger order, and condition order by index finger order, thereby gaining 5 df. When certain variables were linearly dependent on preceding ones (e.g., in the case of electrode leads), the within-error matrix became singular, so that univariate tests were performed.

The main effects and interactions among condition order, index finger order, condition and time slice were tested against $\alpha = .05$ (*). T-tests according to Bonferroni were performed for the main effect of expression type, and the interactions of expression type with condition order, index finger order, condition and time slice: $\alpha = .0167$ (**). The main effect of electrode lead was tested against $\alpha = .0033$ (***), along with its interactions with condition order, index finger order, condition and time slice. Interactions of key order with expression type or electrode lead with expression type, possibly embedded in (combinations of) condition order, index finger order, condition and time slice handled $\alpha = .0011$ (****). Interactions involving key order and electrode lead, but not expression type, yielded 225 contrasts, so that $\alpha = .05/225 = 2.22^{-04}$ (*****). Interactions of key order by electrode lead by expression type rendered 675 contrasts, so that α = .05/675 = 7.41^{-05} (******). Only those effects that were better than these chance levels will be reported.

Results of MANOVA on P300 and N400

Figure 7.6, Table 7.2 and Table 7.3 (Appendix) suggest the following conclusions:

(I) *Expression* elicited more negative going deflections than *context*. When *expression* was presented first, amplitudes were more negative than when presented last. Beginning with the right index finger rendered larger negative amplitudes than beginning with the left, except for the comparison between literals and anomalies in *context*, when *context* was presented last. In this case, the difference reversed. Metaphors and anomalies evoked stronger negative shifts than literals, except that literals yielded stronger negative shifts than metaphors when *context* was presented last and the left hand was used first (interaction of index finger order by condition order by condition by expression type). (II) *Expression* raised the largest negative shifts at fronto-central locations, whereas *context* showed the largest negative deflections at Pz (interaction of condition by lead - see also Figure 7.5). (III) *Expression* affected the amplitudes at 305 and 385 ms equally strong, whereas *context* primarily affected the amplitude at 385 ms (interaction of condition by time slice). (IV) For literal expressions, the amplitude at 385 ms was more negative than that at 305 ms at Pz and C3. For metaphors, however, the amplitude at 305 ms was more negative than that at 385 ms, indicating that the difference between literals and metaphors was centrally left and parietally oriented. For metaphors at Fz and anomalies at Pz, larger negative shifts were elicited at 385 ms than at 305 ms, suggesting that differences between metaphors and anomalies were frontally and parietally oriented (interaction of electrode lead by expression type by time slice).

Figure 7.6: Grand mean amplitudes for literal expressions (■), metaphors (▲) and anomalies (●) at 305 ms (solid) and 385 ms (dashed) after *B-term* onset. O2 rendered no significant effects. Negativity up.

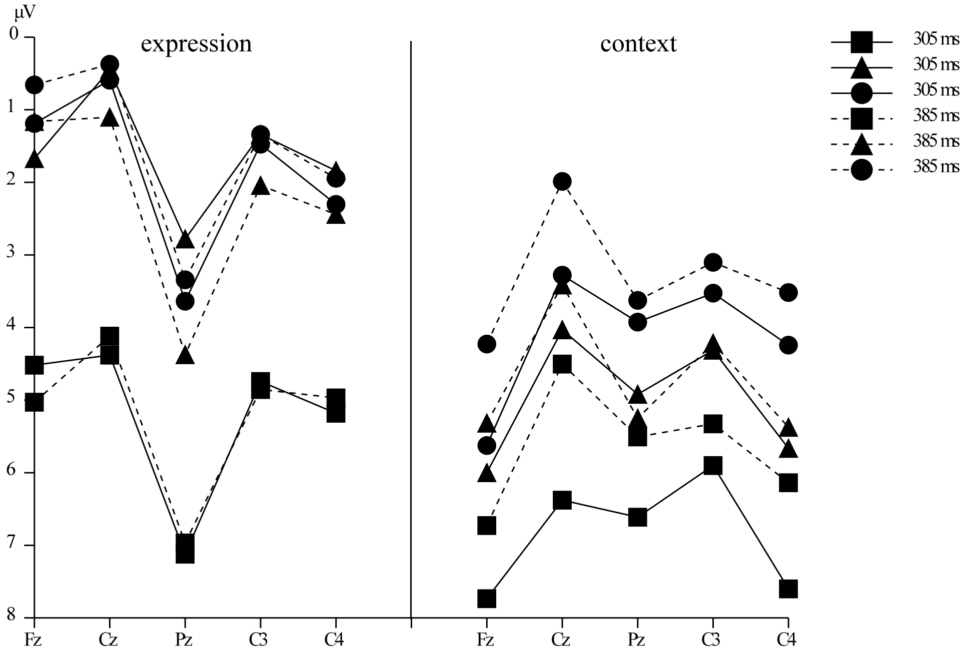


Table 7.2: Grand mean amplitudes (µV) for the interaction of index finger order by condition order by condition by expression type.

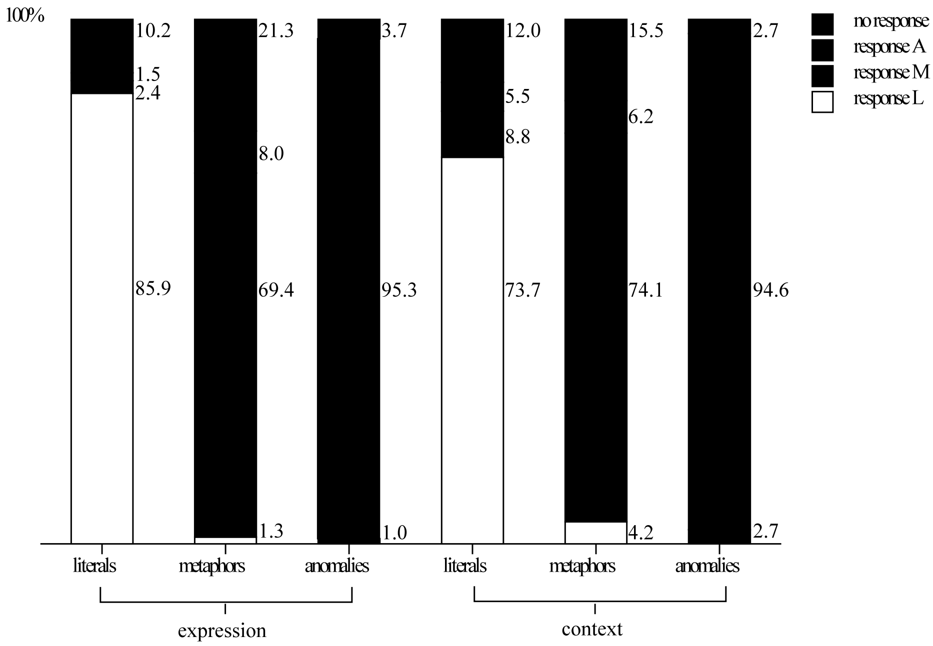
expression before *context*			
right before left finger			
		expression	*context*
	literals	2.08	5.53
	metaphors	1.17	4.13
	anomalies	1.45	2.80
left before right finger			
		expression	*context*
	literals	4.34	4.56
	metaphors	-.36	5.67
	anomalies	-.24	3.41
context before *expression*			
right before left finger			
		expression	*context*
	literals	6.12	5.05
	metaphors	2.14	4.76
	anomalies	1.84	2.67
left before right finger			
		expression	*context*
	literals	8.68	9.61
	metaphors	5.62	5.22
	anomalies	5.39	6.34

Results for decision accuracy

Figure 7.7 shows the proportions of (in)correct decisions in a 3-choice task in *expression* and *context*. Table 7.4 (Appendix) shows the measure of agreement (Cohen's Kappa) of each subject for the correct decisions, the means of which were .762 for *expression* and .726 for *context*.

(I) Anomalies were recognized best throughout conditions (about 95%) and errors tended to 'metaphor'. (II) Metaphors were recognized slightly better in *context* (74.1%) compared with *expression* (69.4%), whereas the opposite occurred for literals: 85.9% in *expression* and 73.7% in *context*. However, MANOVA on the number of correct decisions indicated that the interaction between condition and literals versus metaphors was insignificant (Table 7.4). (III) In *context*, literals tended to be judged more as metaphors.

Figure 7.7: Percentages of (in)correct decisions in a 3-choice task (N = 24).



Discussion of arrival times

Analogous to the RT experiments in Chapter 6, repetition effects affected response times. When *expression* was presented last, RT_a was fastest. However, the differences among expression types remained unchanged. This is different from Chapter 6, where these differences tended to vanish, presumably due to repetition. Similar to the findings in the previous Chapter, anomalies were processed fastest, metaphors slowest, whereas literal expressions were in between. *Context* did not change this pattern. The only difference with the earlier RT_a results is that the trend was significant that literal expressions were reacted to faster than metaphors.

Therefore, **the parallel anomaly model may be maintained**, stating that anomalies are processed in two short stages, literals in a long literal stage and a short figurative one, and metaphors in a long literal and a long figurative stage. It should be noted, however, that repetition of expression types limits this model to the first presentation. Subjects probably recognized earlier judgements, and disregarded the special metaphor processes.

Discussion of the N400

The grand mean EEG showed negative deflections with a peak amplitude at 385 ms after presentation of the *B-term*, which were identified as the N400. Anomalies evoked the highest N400 amplitude, although the difference with metaphors was insignificant. Literal expressions elicited the lowest amplitudes.

It turned out, however, that these differences may change as a function of complex task manipulations. For instance, literals in *context* evoked more negativity than metaphors when *context* was presented last, while the left index finger was used first. Any explanation of this result is merely a wild guess.

Furthermore, it seemed that expression types may be differentiated according to the location of processing. The difference between literals and metaphors took place most prominently at central left and parietal locations, whereas the difference between metaphors and anomalies was most prominent at frontal and parietal locations. Danesi (1989) suggested that processing the *B-term* of a metaphor - as a novel unexpected completion - is a right-hemispheric activity, whereas the *A-term* is processed in the left hemisphere. From this idea, it was inferred that literal *B-terms* - not being novel completions - would be processed left-hemispheric, just as the *A-terms*. The present data show that metaphor processing was obviously not right-hemispheric. No effects that were specific for C4 occurred. Left-hemispheric processing of the literal *B-terms* obtained some support, in that C3 was involved in distinguishing literals from metaphors.

In comparison with other expression types, metaphors affected the moment at which the negative amplitude shifts were strongest. A comparison of metaphors with literals showed that metaphors had their negative peak at 305 ms after *B-term* onset, whereas literals did at 385 ms. Metaphors and anomalies, however, both showed the highest negative shifts at 385 ms. This may suggest that metaphors are earlier separated from literal expressions than from anomalies. In other words, metaphors remain similar to anomalies for a longer period than they do to literals.

The amplitude at 305 ms after *B-term* onset was probably not the reflection of the P300, but of the rising flank of the N400. The rationale is that at Pz and C3, the amplitude at 305 ms for metaphors was even more negative going than at 385 ms, whereas the P300 is supposed to be a positive shift.

Except for the anomaly model, none of the models mentioned different electrocortical effects for the diverse expression types. **As expected by the anomaly model, literal expressions yielded least N400 activity, whereas anomalies yielded most. Since metaphors were less different from anomalies than from literals, it could be concluded that metaphors were intially processed as anomalies, which is a straightforward prediction of the anomaly model**.

However, note that the ordinal pattern of RT_a (A < L < M) is not congruent with the ordinal pattern of the N400 amplitude (A $\geq$ M > L). Thus, **response speed as a function of calculating more feature overlap was not related to higher amplitudes. It may be reasoned that N400 is not part of the processing stages. Probably, it is an additional effect based on a mismatch of semantic categories**.

This, of course, is not a prediction of the anomaly model, which presumes that the electrocortical effect is a functional aspect of the process. In the serial version of the anomaly model, the electrocortical effect was supposed to mark the commencement of the figurative stage. Based on the RT_a results, this is obviously not the case.

In the parallel version, N400 may only have an initial 'signaling' function. In the case of literal expressions, the *A-term* is an instance of the *B-term* category ('their house is a premises'), so that no mismatch occurs. In the case of metaphors ('their house is a grave') or anomalies ('their house is a jerk'), the *A-term* is not an instance of the *B-term* category. Therefore, the *A-term* cannot be integrated into the *B-term* feature set. Here, the N400 effect occurs, which might signal that 'this is probably not a literal expression'. What tips the scale, however, is the success in calculating literal and figurative shared sets, which is a time-consuming control process, launched after the N400-effect.

Eventually, metaphors did not resemble anomalies in the way they were processed. Metaphors established large literal and figurative shared sets, whereas anomalies did not. Therefore, anomalies were processed faster than metaphors. However, metaphors did resemble anomalies in the sense that both provided a category mismatch, evoking similar N400 amplitudes.

The (poetic) context reduced the differences among expression types. It might be that poetry is such a diffuse primer that subjects did not know which stimulus category to expect for the *B-term*, so that the expression types became exchangeable.

It is not likely that N400 was evoked by lower word frequencies for the anomalous *B-terms*, because the metaphoric *B-terms* - which had high word frequencies - elicited similar amplitudes, whereas the literal *B-terms* - with high word frequencies - did not.

Discussion of decision accuracy

Parallel to the results in Chapter 6, error rates were quite high, except for anomalies. For literals in *context*, decisions tended more to 'metaphor' and for metaphors more to 'literal' than compared with *expression*. However, this interaction remained insignificant (Appendix Table 7.4). Despite the extra set of 20 unrepeated practice trials for each expression type in both conditions, the confusion was not resolved.

7.5 **General discussion for psychology**

Fischler in Rugg, Kok, Barrett & Fischler (1985) has already suggested that ERPs may be used to explore metaphor processing. She speculated that ERPs could distinguish literal from metaphoric language, which appears to be valid for the N400. Yet, N400 probably was not part of the metaphor process, but rather the indication of a category mismatch, not present for literal expressions. Some support for this interpretation of the N400 may be found in Boddy & Weinberg (1981), who made an inquiry into the N1-P2 complex in response to category mismatches. Upon visual inspection, their Figure 1 and 2 show larger N400 amplitudes for negative instances of a category than for positive instances. Furthermore, Polich (1985) reported that while subjects performed a semantic categorization task, inappropriate instance-category combinations in a sentence elicited larger N400 amplitudes than appropriate ones.

The idea of category mismatch also converges with the indexing status attributed to N400 by Chwilla, Brown & Hagoort (1995). They saw N400 as a reflection of lexical integration (do words match?), which is congruent with the idea of matching categories. The N400-effect does not seem to be part of the special metaphor process, so that its manifestation may be limited to the less complex forms of semantic processing.

Moreover, various task-induced effects violated the clear-cut semantic relatedness of the N400. Effects of index finger order and condition order intruded upon the N400-effects of condition and expression type. Quite powerful trends of decision key order could only be rejected by handling an extremely conservative α-level. Chwilla, Brown & Hagoort claimed that such side-effects of task requirements can be avoided by installing a passive reading task, which obtained similar results to their lexical decision task. However, this is a trade-off between one dilemma and another. Certainly, task effects are excluded from the data, but all controls over decision errors and constant attention are also lost. This may perhaps not be a problem for a relatively simple task such as lexical decision - indeed, the authors provided evidence for comparable results - but for a complex decision among literals, metaphors and anomalies, errors do frequently occur, so that ERPs should not be calculated across all exemplars of a stimulus category. In addition, by not employing a task in an experiment, the danger of circularity is ever-present. In that case, N400 is not only the reflection of perceiving a semantic unexpectedness, but the unexpectedness is deduced from the manifestation of the N400.

7.6 General discussion for the theory of literature

and he, whom the fire of the spirit	en hem, wien het vuur van den geest
with the bite of a scorpion	met den beet van een schorpioen
pierced the smooth labyrinth of the cortex	door het glad labyrinth van de schors
into the weak of the brain	in het weke der hersenen drong
as poison in a greedy sponge,	als gif in een gulzige spons,
breaks by night in a crown of thorns	breekt bij nacht in een doornenkroon
into the sweat of thought	het zweet der gedachten uit

(Own translation)

Marsman, H. (1975). De dierenriem, VI. *Verzamelde gedichten*. Querido, Amsterdam, 149.

This fragment is one of the contexts used to explore metaphor processing. It speaks about 'the fire of spirit, piercing the cortex, into the brain', and with ample poetic license, it may remind us of the words by Breton that poetry - and specifically metaphor - creates that 'higher unity of fire' or 'the energy of the shot' (Richards 1965: 125), provoked by the collision of semantically deviant words.

The anomaly theory of metaphor processing started exactly from this point, subscribing the surrealist proclamation (Breton 1955) that any two things put on a par make good poetry, no matter how remote or trivial they are. They may not have any symbolic value in themselves, yet, by stating that they *are like* each other, any two randomly chosen objects become art. They form a formal opposition that evokes an effect not present when they are presented in isolation:

> (...) tout [*fait image*] et que le moindre objet, auquel n'est pas assigné un rôle symbolique particulier, est susceptible de figurer n'importe quoi. L'esprit est d'une merveilleuse promptitude à saisir le plus faible rapport qui peut exister entre deux objets pris au hasard et les poètes savent qu'ils peuvent toujours, sans crainte de tromper, dire de l'un qu'il est *comme* l'autre: la seule hiérarchie qu'on puisse établir des poètes ne peut même reposer que sur le plus ou moins de liberté dont ils ont fait preuve à cet égard. (...) Comparer deux objets aussi éloignés que possible l'un de l'autre, ou, par toute autre méthode, les mettre en présence d'une manière brusque et saisissante, demeure la tâche la plus haute à laquelle la poésie puisse prétendre. En cela doit tendre de plus en plus à s'exercer son pouvoir inégalable, unique, qui est de faire apparaître l'unité concrète des deux termes mis en rapport

> et de comuniquer à chacun d'eux, quel qu'il soit, une vigueur qui lui manquait tant qu'il était pris isolément. Ce qu'il s'agit de briser, c'est l'opposition toute formelle de ces deux termes. (...) Plus l'élément de dissemblance immédiate paraît fort, plus il doit être surmonté te nié. (...) Ainsi deux corps différents, frottés l'un contre l'autre, atteignent, par l'étincelle, à leur unité suprême dans le feu. (Breton 1955: 148-149)

It is this 'higher unity of fire' that was investigated in the current Chapter. It was supposed that metaphors and anomalies would evoke certain electrocortical effects, not present, or at least less powerful for literal expressions.

Indeed, such effects were found. A brain potential called N400 was examined, which - in previous psychological studies - was found to be sensitive to the frustration of semantic expectations. Literal expressions appeared to elicit smaller N400 amplitudes than metaphors and anomalies.

The question posed by Danesi (1989) whether the brain has specialized areas to process metaphors was partly confirmed. Danesi stated that the *B-term* of a metaphor is a novel completion of the *A-term* and that the physiological structure of the right hemisphere was most suitable to process novel stimuli. It was deduced that normal completions, such as literal *B-terms*, would be processed by the left hemisphere, which - as Danesi suggested - was better equipped to process overlearned language patterns. However, metaphor processing was not a right-hemispheric activity. No N400-effects occurred for expression type that were specific for the electrodes put at the right side of the skull. More evidence was found for the idea that literal *B-terms* were processed by the left hemisphere. The distinction between literals and metaphors took place particularly at the electrodes placed on the left side of the skull.

Yet, it was inferred that the N400-effect was not part of the processing of the expressions, but that it was an additional effect based on a category mismatch. The reason was a dissociation between the pattern of RT_as and the pattern of the N400. The RT_as showed that anomalies were fastest, followed by literals, which were faster than metaphors. If a brain potential is part of a process, it is supposed to follow the pattern of RTs. Thus, N400 should have been largest for anomalies, followed by literals, and smallest for metaphors, which was obviously not the case.

A feasible alternative was that metaphors and anomalies provided a mismatch between the category of the *A-* and the *B-term*, which was not present for literal expressions. Literal expressions matched an instance with an appropriate category ('their house is a premises'), whereas metaphors ('their house is a grave') and anomalies ('their house is a jerk') did not (cf. *Instance-category distribution per expression type*, Chapter 4).

In sum, Breton's idea that a formal opposition is formed by nonliteral expressions was correct in that metaphors and anomalies both contain a category mismatch, which is probably responsible for the strong N400-effects. In this respect, the anomaly model, which states that metaphors and anomalies resemble each other is also correct, in evoking a larger electrocortical effect than literal expressions. However, this effect is not an inalienable part of the process that distinguishes metaphors from other expression types (cf. the RT_a pattern), but is rather an additional category mismatch effect. This, of course, was not foreseen by the anomaly model.

In other words, if N400 is the reflection of a formal opposition in semantics - such as a category mismatch - Barthes' question - whether the structuralist principle of opposition should be reduced to terms of binarism - becomes most germane. Barthes envisioned the communication in binary codes as a general human phenomenon, thereby referring to the drum codes of tribal man, as well as the automatic digital information processing and cybernetics. According to Barthes, binarism expresses itself most strongly in artificial systems and is less profound in natural communication. Barthes wondered if binarism is a universal phenomenon; and on the other hand, given it's universality, if it has a natural foundation:

> Questioned in phonology, unexplored in semantics, binarism is the great unknown in semiology, whose types of opposition have not been outlined. (...) It is very tempting to found the general binarism of the codes on physiological data, inasmuch as it is likely that neuro-cerebral perception also functions in an all-or-nothing way (...). (Barthes 1970: 82)

The latter remark is not really pertinent. Neuro-cerebral perception does function in an all-or-nothing way at the level of the cells. At the synaps, electrons are either fired or not. Certain cells even await a preset number of signals before starting to fire, which, in a way, may be seen as a binary procedure. However, this is not the level that Barthes wanted to study, because the semantic oppositions occur at the level of cognitive processing.

Here, the concept of binary systems is not so self-evident. In Donders (1868) and in the earlier work of Sternberg (1969), the supposition was that information accumulated within a stage, before it was transmitted to the next stage in an all-or-nothing way. By contrast, later models allowed a gradual build up and constant flow of information between stages (e.g., McClelland 1979). Ratcliff (1988) proposed that decisions between two options (e.g., is the word pair an opposition or a synonym?) are liable to gradual stochastic mutations in strength among the various response options. These latter two theories are definitely not binary models. Since the stage output is not 'on' or 'off', the time to execute a process is not the sum of its stages, as in a binary model.

It may even be questioned whether the oppositions and semiological codes are binary at all. A binary code knows no overlap between the onset or offset of an electric switch. The opposition in a binary code is formed by the presence or absence of a signal. Oppositions in meaning are not the presence or absence of one word or code, but the presence of two words or codes with strong distinctive meanings. Yet, because oppositions emerge from the comparison of two meanings, they hardly ever come without semantic overlap. In other words, the biological signal is analogue rather than binary.

What emerges is that metaphors combine literal and figurative features in the shared set. This is the most time-consuming operation compared with literal expressions, which share literal features, and anomalies which share hardly any features. Metaphors and anomalies evoke special electrocortical effects, but these are probably established by the category mismatch that is inherent in these expression types, not by the size of the shared feature set. These findings agree best with a parallel anomaly model, which will be elaborated in the next, concluding Chapter.

Notes:

1. However, the results cannot be interpreted as straightforwardly as suggested by Winner & Gardner. Instead of analyzing all subject groups and factors in one analysis, subsets of subjects and factors were selected, thereby discarding the inherent interactions, and unintentionally increasing the degrees of freedom. In the original design, eight groups of subjects were tested: 10 normal controls, 4 left-anterior aphasics (Broca), 10 left-anterior aphasics (no Broca), 7 left-posterior aphasics (Wernicke), 6 left-posterior aphasics (no Wernicke), 8 patients with anterior and posterior lesions, 7 demented patients, 22 right-hemisphere patients.

These subjects were confronted with 18 adjective-noun metaphors ('a heavy heart'), half of which were synaesthetic ('colourful music'). Although the aim was to investigate left- and right-hemisphere sensitivity for literal versus figurative language, no contrast was provided with literal expressions. According to the authors, the results for the two metaphor types did not differ, and were pooled, without, however, providing the statistics that grounded this decision.

For each expression, four pictures served as interpretation of the metaphors. They could depict the metaphoric meaning, the literal meaning, the adjective or the noun. Actually, the latter two were not interpretations, but rather direct pictorial representations of one word, the recognition of which is not comparable with interpreting a metaphor.

Subjects chose which picture best expressed the meaning of the metaphors, and two scores were obtained: The number of first and the number of second choices. Although nominal data demand a χ^2 statistics, analysis of variance was performed. Instead of one overall analysis, separate analyses were performed: A ONEWAY analysis of variance for the four pictures in four subject groups, a t-test (expressed in F-values) for left- versus right-hemisphere patients, and an analysis of variance on right-hemisphere patients versus anterior versus posterior aphasics. The complete design, however, was an 8*2*4*2 MANOVA of subject group by metaphor type by picture type by score type. Moreover, it is doubtful whether different receptiveness between left- and right-hemisphere patients was measured for metaphor understanding, or that right-hemisphere patients had more difficulty in transforming a linguistic text into a pictorial representation.

2. Repetition effects on the N400 were reported by Rugg (1985, 1987), Rugg & Nagy (1987), Rugg, Furda & Lorist (1988), Nagy & Rugg (1989), Roth & Boddy (1989).

3. The frequency domain is a decomposition of sinus waves of different frequencies. In other words, a time series is analyzed as the frequencies $\{0, \ldots., n\}$, such that $\sum(\omega)$ results into the original time series. In the frequency domain, however, time does not exist, only frequencies.

4. Note that it is common practice in encephalography to reverse the polarity in the presentation of the data. Therefore, negativity is up.

Appendix to Chapter 7

Table 7.0: MANOVAs for the effects of condition order, condition and expression type on the grand mean arrival time (RT_a) for correct responses. Insignificant effects are not tabulated.

Interactions of condition order by condition:

1 = ((*expression* before *context*) vs (*context* before *expression*)) vs
(*expression* vs *context*)

dep. variable	$F_{1,22}$	p	df	quasi-F	p		coeff.	t	p
RT_a	17.52	.000	$_{1,24}$	15.80	.001	1	244.46	4.19	.000*

Main effects of condition:

1 = *expression* vs *context*

dep. variable	$F_{1,22}$	p	df	quasi-F	p		coeff.	t	p
RT_a	26.72	.000	$_{1,22}$	26.81	.000	1	-301.93	-5.17	.000*

Main effects of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,21}$	p	df	quasi-F	p		coeff.	t	p
RT_a	.84	53.45	.000	$_{2,30}$	41.12	.000	1	-157.71	-4.13	.000**
							2	225.70	7.73	.000**
							3	383.41	9.75	.000**

Table 7.1: F-values for the minimal and maximal amplitude differences at 10 intervals between 265 and 445 ms after *B-term* onset. Temporal locations are neglected, due to extensive numbers of outliers. N = 24.

Main effect of condition:

expression vs *context*

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	17.44	.0004**	27.35	.0000**
Cz	1.31	.2634	22.29	.0001**
Pz	.00	.9609	.99	.3312
Oz	.46	.5021	3.62	.0697
C3	3.47	.0752	13.18	.0014**
C4	6.91	.0150*	27.67	.0000**

Main effect of expression type:

literals vs metaphors

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	11.36	.0026**	28.34	.0000**
Cz	1.34	.2588	24.34	.0001**
Pz	.00	.9632	25.28	.0000**
Oz	.27	.6104	10.70	.0034**
C3	2.65	.1170	26.05	.0000**
C4	1.38	.2523	25.52	.0000**

literals vs anomalies

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	13.06	.0015**	29.58	.0000**
Cz	4.44	.0461*	24.33	.0001**
Pz	2.54	.1244	25.37	.0000**
Oz	.20	.6598	21.33	.0001**
C3	7.09	.0139*	27.01	.0000**
C4	6.42	.0186*	34.25	.0000**

metaphors vs anomalies

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	.53	.4753	1.78	.1948
Cz	.23	.6354	2.00	.1707
Pz	.01	.9042	3.45	.0761
Oz	.04	.8371	3.38	.0790
C3	.42	.5224	3.18	.0877
C4	.91	.3493	3.59	.0707

Interaction of expression type by condition:

literals (*expression* - *context*) vs metaphors (*expression* - *context*)

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	1.06	.3145	5.70	.0256*
Cz	1.51	.2315	2.92	.1009
Pz	.96	.3371	6.60	.0172*
Oz	.60	.4449	3.78	.0643
C3	1.98	.1728	9.42	.0054*
C4	.94	.3417	5.12	.0334*

literals (*expression* - *context*) vs anomalies (*expression* - *context*)

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	1.03	.3211	2.44	.1316
Cz	.32	.5780	1.67	.2086
Pz	.14	.7128	4.92	.0367*
Oz	.00	.9636	1.08	.3101
C3	.62	.4403	3.97	.0582
C4	.00	.9646	.47	.5016

metaphors (*expression* - *context*) vs anomalies (*expression* - *context*)

electrode location	min. amp. diff. $F_{1,23}$	p	max. amp. diff. $F_{1,23}$	p
Fz	.00	.9734	1.67	.2094
Cz	.28	.6040	2.08	.1628
Pz	.02	.8924	2.47	.1294
Oz	.71	.4093	4.45	.0461*
C3	.08	.7800	2.41	.1340
C4	2.83	.1062	9.48	.0053*

Table 7.3: MANOVA for the effects of condition order (*expression* before *context* vs *context* before *expression*), index finger order (left before right vs right before left), lead (Fz vs Cz vs Pz vs Oz vs C3 vs C4), condition (*expression* vs *context*), expression type (literal vs metaphor vs anomaly) and time slice (305 ms vs 385 ms). Effects of decision key order were never better than a conservative α-level.

Main effect of condition:

1 = *expression* vs *context*

dep. variable	$F_{1,5}$	p		coeff.	t	p
μV	16.81	.009	1	-64.46	-4.10	.009*

Main effect of expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,4}$	p		coeff.	t	p
μV	.96	54.04	.001	1	-52.85	11.20	.000**
				2	66.98	6.21	.002**

Interaction of condition order by condition:

dep. variable	$F_{1,5}$	p	coeff.	t	p
μV	6.93	.046	-41.37	-2.63	.046*

Interaction of index finger order by condition order by condition by expression type:

1 = literals vs metaphors
2 = literals vs anomalies
3 = metaphors vs anomalies

dep. variable	Pillai's Trace	$F_{2,4}$	p		coeff.	t	p
μV	.78	7.04	.049	1	-33.98	-3.58	.016**
				2	-22.24	-3.85	.012**

Interaction of condition by electrode lead:

1 = (*expression* vs *context*) vs (Fz vs Pz)
2 = (*expression* vs *context*) vs (Cz vs Pz)
3 = (*expression* vs *context*) vs (Pz vs C4)

dep. variable	Pillai's Trace	$F_{5,1}$	p		coeff.	t	p
μV	1.00	419.85	.037	1	-19.78	-6.39	.001***

dep. variable	$F_{1,5}$	p		coeff.	t	p
μV	102.66	.000	2	-10.96	-10.13	.000***
	31.93	.002	3	12.24	2.65	.002***

Interaction of condition by time slice:

1 = 305 ms vs 385 ms

dep. variable	$F_{1,5}$	p		coeff.	t	p
µV	15.33	.011	1	-14.02	-3.92	.011*

Interaction of electrode lead by expression type by time slice:

1 = (Pz vs C3) vs (literals vs metaphors) vs (305 ms vs 385 ms)
2 = (Fz vs Pz) vs (metaphors vs anomalies) vs (305 ms vs 385 ms)

dep. variable	$F_{1,5}$	p		coeff.	t	p
µV	42.65	.001	1	2.11	6.53	.001****
	73.16	.000	2	1.78	8.55	.000****

Table 7.4: Cohen's Kappa and MANOVA for the correct decisions of each subject in a 3-choice task in *expression* and *context*.

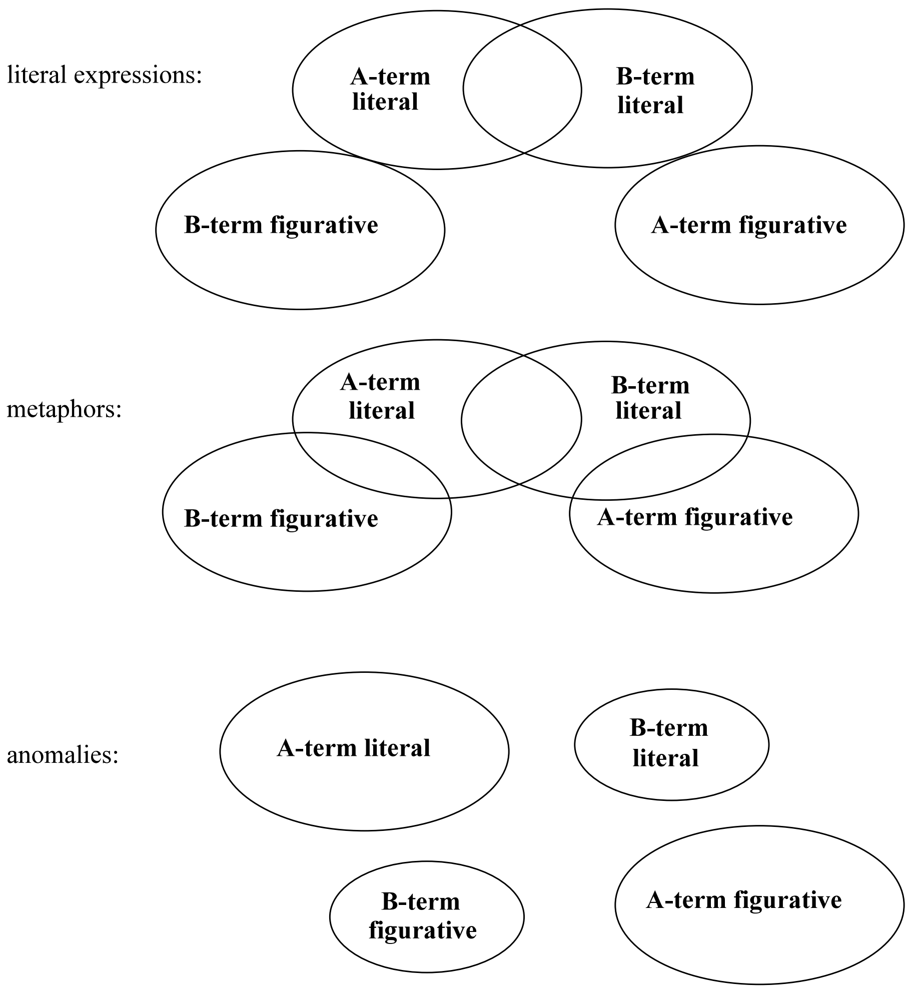
	expr.	***context***		***expr.***	***context***	
subject	**κ**	**κ**	**subject**	**κ**	**κ**	
01	.70	.55	13	.65	.56	
02	.57	.79	14	.79	.81	**interaction of condition by**
03	1.00	.84	15	.96	.81	**expression type:**
04	.71	.65	16	.71	.61	Pillai's Trace = .32
05	.80	.83	17	.86	.68	$F_{2,22}$ = 5.23, p = .014
06	.58	.88	18	.84	.93	
07	.80	.77	19	.80	.59	**cond. by (literals vs metaphors):**
08	.85	.72	20	.86	.76	coeff. = 3.16, t = 1.88, p = .073
09	.74	.80	21	.62	.58	**cond. by (literals vs anomalies):**
10	.76	.64	22	.70	.81	coeff. = 3.33, t = 3.29, p = .003
11	.55	.81	23	.86	.52	**cond. by (metap. vs anomalies):**
12	.81	.76	24	.89	.79	coeff. = .17, t = .12, p = .909
$\bar{x}$				**.762**	**.726**	

CONCLUSIONS

The parallel two-stage anomaly race model

Three principal findings form the underpinning of the new anomaly model. First, metaphors accumulate a literal and (mixed) figurative shared set. Literal expressions accumulate only a literal shared set,

Figure 8.0: Parallel anomaly model for a four sets feature overlap in literal, metaphoric and anomalous expressions.



and anomalies hardly any shared set at all. Second, the time for accumulating the shared set is shortest for anomalies, whereas accumulating the shared set for metaphors takes a time that is equal to or longer than for literal expressions. The reason is that anomalies activate smaller numbers of features and accumulate small shared sets, so that processing stops quickly. For literals and metaphors, more features are activated and larger shared sets are accumulated, so that processing time increases. Thus, the time to decide among expression types not only depends on the serial order of criterion checks for sufficient overlap, but primarily on an 'end of file' criterion. The third finding is that metaphors and anomalies evoke a more negative N400 amplitude than literal expressions, which may be ascribed to category mismatches for the former two.

Figure 8.0 shows the Venn diagrams for the feature sets of literal, metaphoric and anomalous expressions, according to the parallel anomaly model. Only functionally important shared sets are depicted. Literal expressions accumulate large literal shared sets. Metaphors accumulate shared sets consisting of literal and figurative features as well. Anomalies do not accumulate a considerable number of shared features. *Context* does not seriously affect this configuration. Unlike Figure 2.0, Chapter 2, Figure 8.0 suggests that the *A*- and *B-terms* activate equal set sizes, except for anomalous *B-terms*, which have considerably smaller shared sets.

Figure 8.1 shows a parallel anomaly model. The activation of features in the *encode phase* is similar to that in the comparison model (Figure 1.1, Chapter 1). The *A-term* activates feature set *X* and the *B-term* feature set *Y*. The features are words, rather than lines or letters. In this phase, it is unimportant whether features are literal or figurative, and in which order they arrive. The chance that the first feature is literal may be high, but figurative features intermingle. Yet, the features are already indexed as literal and/or figurative: $X = X_l \cup X_f$ and $Y = Y_l \cup Y_f$. This is quite different from the serial anomaly model (Figure 2.1, Chapter 2), in which literal features were activated and processed before the figurative features.

Moreover, a check for category membership is introduced in the *encode phase*, which may lead to the N400-effect that accompanies 'lexical integration' (Chwilla, Brown, & Hagoort 1995). After the *B-term* generated feature set *Y*, it is verified whether the *A-term* is an instance of set *Y*. (Evidently, it may also be checked whether the *B-term* is part of *X*). In other words, it is assumed that appropriate instance-category combinations are more likely to be activated by each other than inappropriate ones. In the case of literal expressions, the *A-term* is an appropriate feature of *Y*, so that no or less N400-activity occurs. This may lead to a facilitation for calculating the literal shared set S_l, in that the route to arrive at a decision for 'literal' is elevated to a higher level of activation than the route for 'metaphor' or 'anomaly'. If the *A-term* is not a part of *Y*, the category mismatch may evoke N400-activity and facilitate the calculation of the (mixed) figurative shared set S_f. In that case, the route to arrive at a decision for 'metaphor' or 'anomaly' is activated more strongly than the route for 'literal'. Due to the category mismatch, N400 would form a preliminary indication whether the expression is 'literal or not'.

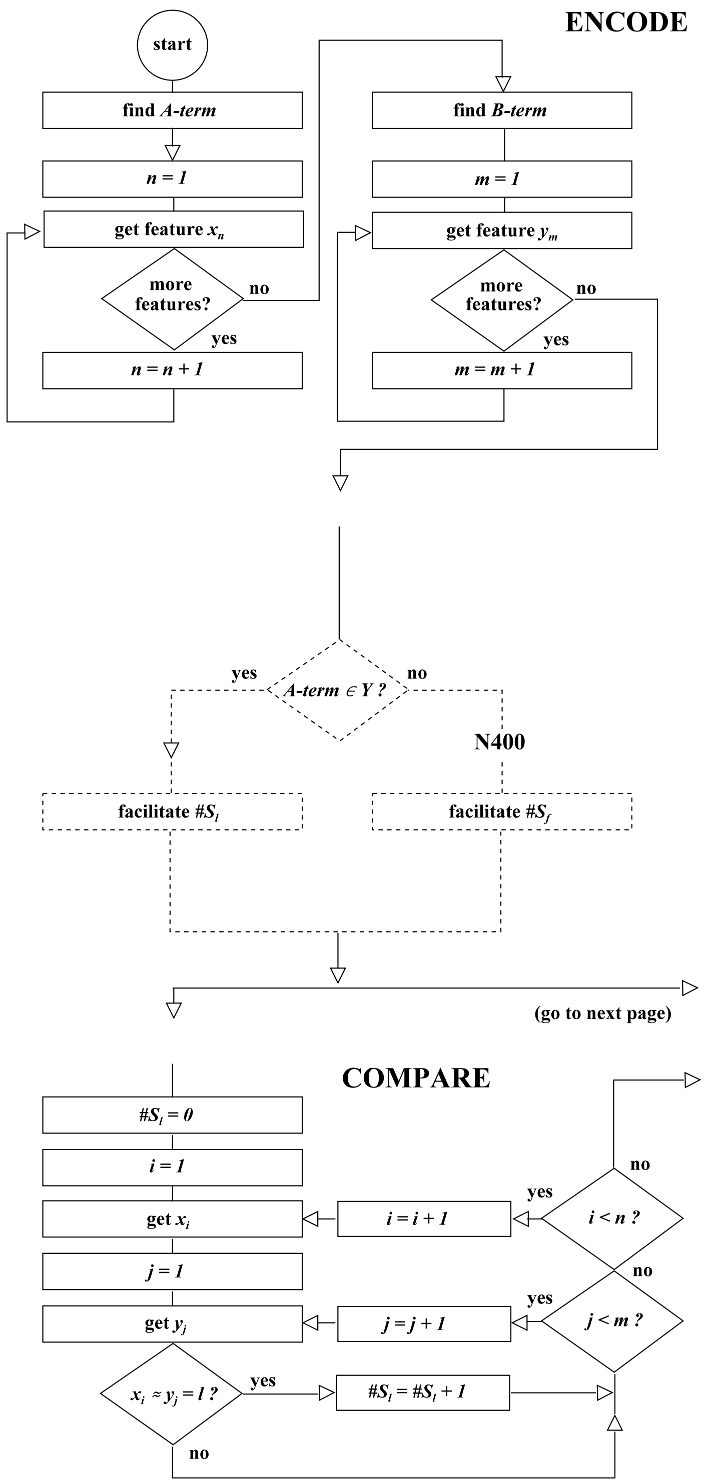
Beware, however, that with the facilitation of the literal or figurative route, N400 is attributed a functional role in the metaphor process, which may in fact not be present. Since the status of this process module is questionable, it is represented by broken lines.

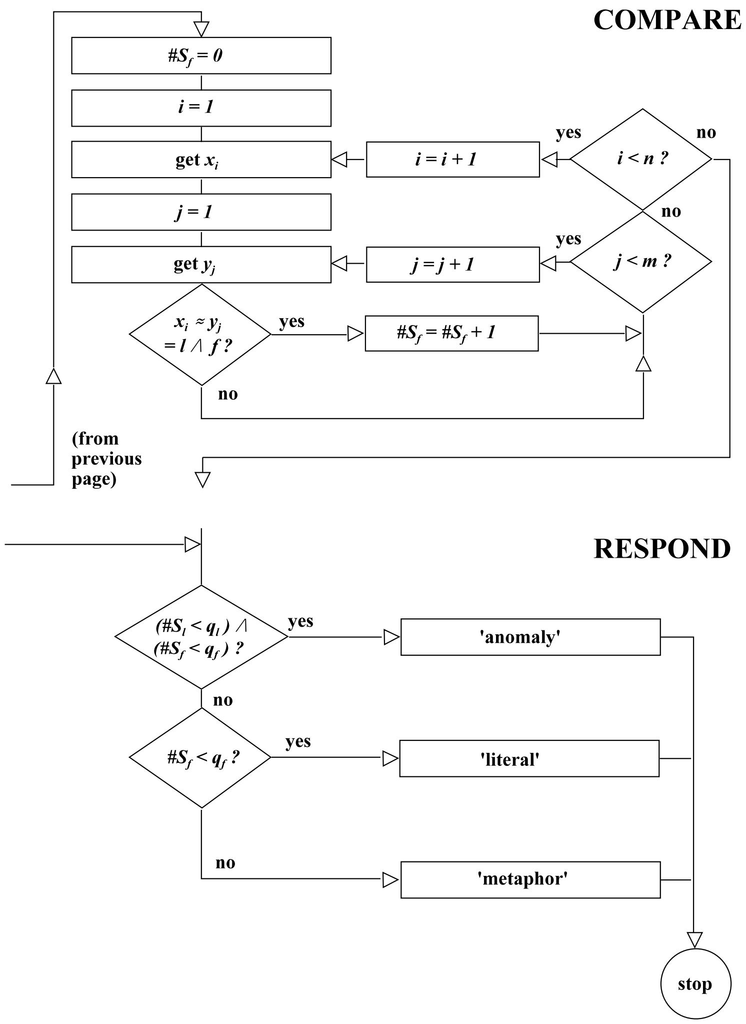
In the *comparison phase*, two information accumulators ($\#S_l$ and $\#S_f$) operate in parallel, and race each other to satisfy the response criteria. Each feature x_i is compared with each feature y_j , using an intermediate criterion of equality ($x_i \approx y_j$?). Thus, derivations, inflections, singular and plural are counted as equal to each other and to word stems. For the accumulation of $\#S_l$, only those shared features that are literal ($x_i \approx y_j = l$?) are counted. Meanwhile, the calculation of $\#S_f$ asks whether a shared feature is literal for the one set and figurative for the other: $x_i \approx y_j = l \wedge f$? . With this procedure, shared features that are figurative for both feature sets are filtered.

The information accumulator that reaches 'end of file' first is the first to enter the *response phase*. Anomalies will be fast for both accumulators, because smaller feature sets are activated by the *B-term* and few shared features are found. Literal expressions may be fast for $\#S_f$ and slower for $\#S_l$, for which many literal shared features are found. Metaphors are slow for both $\#S_l$ and $\#S_f$, because they establish literal as well as (mixed) figurative shared sets. Yet, the *comparison phase* allows that the race between $\#S_l$ and $\#S_f$ ends in a draw, so that the high speed of processing anomalies should be regulated by the *response phase*.

The *response phase* is a combination of serial and parallel criterion checks. The check for 'anomaly' precedes that for 'literal' and 'metaphor'. However, the decision for anomality inspects two criteria in parallel. A minimum number of literal shared features should be found along with a minimum number of (mixed) figurative shared features before an expression can be considered literal or metaphoric. The criterion q_l is the number of literal shared features needed to become 'literal' and corresponds with the upper panel of Figure 8.0. The criterion q_f is the number of (mixed) figurative shared features needed to become a metaphor, and corresponds with the middle panel of Figure 8.0. If an expression satisfies q_l nor q_f , it is an anomaly: $(\#S_l < q_l) \wedge (\#S_f < q_f)$? . If an expression exceeds q_l or q_f , it only needs to be checked whether sufficient (mixed) figurative shared features have been accumulated. If $\#S_f$ is smaller than q_f , $(\#S_f < q_f)$? , the expression is literal. If q_f is met or exceeded, the expression is a metaphor.

Figure 8.1: Parallel two-stage anomaly race model for metaphor understanding during the first encounter in *expression* or *context*.





Modeled in this way, the *response phase* allows for a continuum from literalness to metaphoricity to anomality (cf. note 5, Chapter 1). Suppose that the critical values of q_l and q_f are 10 shared features each. If an expression accumulates 9 shared features for both $\#S_l$ and $\#S_f$, the expression is an anomaly. If $\#S_l$ is 9 features, while $\#S_f$ is 100, the expression is a metaphor that tends to anomality. If $\#S_l$ is 100, while $\#S_f$ is 9, the expression is literal. If $\#S_l$ is 100, while $\#S_f$ is 11, it is a metaphor with a literal trend. In other words, there may be fast metaphors (0 for $\#S_l$, 10 for $\#S_f$) and slow literals (100 for $\#S_l$, 9 for $\#S_f$). In a categorization task as performed in Table 4.1 (Chapter 4), the expressions with strong trends for other expression types would probably not reach significance.

Implementation of the parallel anomaly model would require the following computational procedures:

1. encode *A-term*: $X = X_l \cup X_f = \{x_1.....x_n\}$
2. encode *B-term*: $Y = Y_l \cup Y_f = \{y_1.....y_m\}$
3. if *A-term* ε $Y \rightarrow$ facilitate $\#S_l$
   else facilitate $\#S_f$
4. compute $\#S_l = (X_l \cap Y_l) \wedge \#S_f = (X_l \cap Y_f) \cup (X_f \cap Y_l)$
   if $(\#S_l < q_l \wedge \#S_f < q_f) \rightarrow$ respond 'anomaly' and stop
   else go to 5
5. if $\#S_f < q_f \rightarrow$ respond 'literal'
   else respond 'metaphor'
6. stop

q_l = number of literal elements in $X_l \cap Y_l$ needed to respond 'literal'
q_f = number of literal and figurative elements in $(X_l \cap Y_f) \cup (X_f \cap Y_l)$ needed to respond 'metaphor'

The model may employ weighted features, which means that certain literal and figurative features of a term may be more important than others (e.g., more salient, more prototypical, more frequently mentioned). The weighing of the features could lead to a preferred order of comparison. In that case, $\#S_l$ should be rewritten as W_{Sl} and $\#S_f$ as W_{Sf}, which is the weighted set size W of all shared features u, or the sum of all weights w of all features u in S_l and S_f (cf. Section 2.3, Chapter 2).

The predictions in accordance with the parallel anomaly view are:

1. **A metaphor activates four feature sets: Each term activates literal and figurative features without a fixed order**
2. **the set sizes of these four feature sets are not always fixed: *Single term* raises more features than *expression* or *context***
3. **A literal expression has more shared literal features than a metaphor and an anomaly; a metaphor has shared features that are literal for the one term and figurative for the other more than a literal and an anomalous expression**
4. **The shared sets are not fixed, and cannot be predicted from the isolated words: *Expression* and *context* raise larger shared sets than *single term***
5. **The response time for understanding an anomaly is less than for literals and metaphors. Literals may have a slight advantage, however, they may end in a draw with metaphors**
6. **Metaphors and anomalies show large electrocortical effects for category mismatches, not or less present for literal expressions**
7. **Between the two, *context* and *expression* usually show no significant**

differences. When they do, *context* nullifies the effects obtained in *expression*

8. **From the second presentation on, expression type differences will be dissolved**

Take into consideration that the parallel anomaly model only applies to the first presentation of expressions. Repetitions of identical expressions wipe out all expression type differences, because subjects probably recognize the stimuli. Thus, the special metaphor processes may be ignored in favor of a simple replication of an earlier judgement.

Moreover, basic linguistic properties such as word frequency and lexical ambiguity may disrupt the process. Different word frequencies and lexical ambiguity may increase the noise of the data. The effects of these variables reiterate the point that stimulus sets should be as consistent as possible. Thus far, and including this study, metaphor research has not streamlined the stimuli sets rigorously enough, so that the results are open to unwanted side-effects. Putting much effort into the stimulus sets is rewarded by a higher resolution of the observed effects (see last Section).

Furthermore, Steen (personal communication) hesitated to accept that the feature elicitation during one minute (Chapter 5) would reflect the feature activation (and comparison) during the first seconds of processing (Chapter 6 and 7). Indeed, this is a major problem in testing the on-line processing of metaphors. At the moment, however, there is no alternative. Therefore, it should be kept in mind that the model in Figure 8.1 is a reconstruction, rather than a finding.

On MIDAS

The parallel anomaly model has certain traits that correspond with an earlier computational approach, called MIDAS (Metaphor Interpretation, Denotation, and Acquisition System: Martin 1992). As in the above proposed model, MIDAS is subject to the 'total time constraint' (Gerrig 1989_a), which means that literal expressions and (conventional) metaphors (e.g., idioms) are processed with roughly equal rapidity.

Unlike the parallel anomaly model, MIDAS works with a semantic network of hierarchically organized concepts. In the parallel anomaly model, two feature sets are compared, and the shared sets determine the expression type. In MIDAS, the search strategy is to ask whether the *B-term* is part of the *A-term* concepts. Thus, MIDAS essentially asks for category membership. For instance, when MIDAS encounters a conventional metaphor such as 'to kill a process', it checks whether the *B-term* ('process') is part of the concepts of the *A-term* ('kill') that are available in the memory base. Such a network of concepts differs from a feature set, in that it already knows the fixed relations among its members. Thus, 'kill' presumes 'killer' and 'kill-victim'. MIDAS checks whether 'process' can perform the same role as 'victim', which is predefined in the semantic network. In this respect, MIDAS more closely resembles an interaction model than an anomaly model.

However, the total time constraint is abandoned when MIDAS encounters a (conventional or innovative) metaphor about which no knowledge is stored. Then, an additional, serial stage is executed in the form of a search for the closest related associate (MES: Metaphor Extension System).

The difficulty with the search for the most closely related associate - which is nice as a problem-solving tool for its own sake - is that novel metaphors may be processed equally quickly as literals, whereas in MIDAS's conception, extra time is needed to execute the MES. The nice thing about the MES is that metaphors stored in the memory base may be used to reason what the solution to an unknown metaphor might be. Suppose that 'to kill a process' is the unknown metaphor, whereas MES knows that 'to kill a conversation' means 'to terminate a conversation'. MES also knows that a process can be terminated. Thus, the match ('terminate') is established by opening the semantic network of an *A-term* concept ('kill conversation') and to compare its concepts with concepts (e.g., 'terminate') in the *B-term* network (the network of 'program'). Such a device is not implemented in the parallel anomaly model; however, if it is allowed that features also activate feature sets, a similar mechanism could be applied.

On the other hand, MIDAS cannot really distinguish literals from metaphors, because it interprets each expression by checking all available concepts, literal or metaphoric. If the memory base is not sufficient, MIDAS presumes that it has encountered a metaphor, which need not be the case. In principal, if MIDAS attempted to interpret a literal expression of which it has no knowledge, it would treat it as a metaphor by starting the MES. In MIDAS, every use of an expression is predefined, without a criterion to decide between literalness or metaphoricity.

MPC: A neural net for metaphors

Another approach to metaphor comprehension is the use of a neural network system. With a neural net, it may be accounted for that subjects learn from repeated presentations, although the sudden disappearance of expression type differences in RT_a from the second presentation on is harder to explain. Neural nets suppose a gradual learning curve, which is described by changing the weights for a stimulus. However, if the sudden disappearance of expression type differences in RT_a is a matter of recognition, then the gradually changing weights loose their explanatory power. The equal RT_as for repeated presentations indicate 'one trial learning', whereas neural nets presume that such extraordinary changes of weights are established by a more gradual knowledge build up. The weakness of a neural net approach is that it has problems accounting for abrupt changes in weights, so that outliers become problematic.

A neural net approach to metaphor was advanced by Thomas & Mareschal (in press). The Metaphor by Pattern Completion (MPC) model conceives of simple '*A* is *B*' metaphors as a semantic vector representation of the *A-term* to a connectionist network storing the knowledge base of the *B-term*. The MPC model accounts for how the semantic features of the *A-term* are transformed by the semantic features of the knowledge base of the *B-term*.

'*A* is *B*' metaphors were formed among three terms: Apples, Balls, and Forks. These were defined by a set of prototypical tokens, representing different kinds of apples, balls, and forks, which functioned as the knowledge base. In the case of 'the Apple is a Ball', the output would be an Apple transformed so as to make it more consistent with the prototypical Ball representation stored in the network. The nature of the transformation depends on the relationship between the Apple and Ball features.

After the network was run for 'the Apple is a Ball', certain features of Apple were weighted less (edibility), whereas others were enhanced (e.g., round and hard). Asymmetric comparisons ('the Ball is an Apple') were handled by reducing or enhancing the weight of other features than for 'Apple is Ball', although the similarity between Apple and Ball - viewed as a distance measure - remained equal. The MPC model was also sensitive to context, in that 'the brown Apple is a Ball' resulted in even smaller weights for 'edibility', and an increase in, for example, 'softness'. However, this happened only under the condition that the relation was specified in the knowledge base that brown apples have different features from red apples, i.e. that they are not edible. Furthermore, the MPC model could account for variants such as 'the Apple is a Baseball', or 'the Apple is a Beachball', enhancing or reducing the weights of the relevant features.

Most of the characteristics of the MPC model may be equally well represented by a conventional feature-matching model. The fact that certain features receive more weight than others is easily established if features in the shared set are weighted heavier than the distinctive features. Context sensitivity is accounted for if the feature sets are enriched with relations among features. For example, when 'soft' is a shared feature between Apple and Ball, it may have a relation ('brown') under Apple that is connected to the feature 'not edible'. Thus, the uneatable status of Ball is increased by the shared feature 'soft', which maintains a relation ('brown') with 'inedibility' for Apple. That other features for 'Apple is Baseball' receive more weight than for 'Apple is Beachball' is also not a problem, because Baseball and Beachball activate different feature sets, with different sets of (perhaps heavier weighted) shared features.

The issue of asymmetric comparisons is somewhat different. The MPC model assumes that 'Apple is Ball' has a different meaning than 'Ball is Apple', because the weights of the shared features change, thus causing the asymmetry. However, this explanation overlooks the confounding that the reversed metaphor also makes a category shift. 'Apple is Ball' actually states that apples belong to the category of balls, and thus, that 'apple' is a feature of Ball. Vice versa, 'Ball is Apple' states that balls are elements of the category of Apples, and thus, that it is a feature of Apple. In other words, it is not only the different weights of the shared features that may play a role in the asymmetry effect, but also the response availability of 'Apple' as an instance of 'Balls' and 'Ball' as an instance of 'Apples'. In other words, the frequency with which one term is mentioned as a feature of the other may determine the crucial difference in meaning.

Moreover, if shared features that belong to the *A-term* (the focus) of the comparison are weighted heavier that those belonging to the *B-term*, a feature matching model may account for the same results. However, as shown in Chapter 5, there was no evidence for asymmetry effects of the (shared) feature sets, so that the matter of asymmetric comparisons may turn out to be purely academic.

The mechanism for making a distinction between metaphors and literal expressions is problematic in the MPC model. The problem is dealt with in the following way. An expression is recognized as literal if only small semantic changes occur in the output. For 'this [apple-like] item is an apple', little change is produced in the weights of the item's features. However, for 'this [apple-like] item is a ball', the weights will change drastically. The criterion that determines the shift from literal expression to metaphor is detected by matching the network error score against a value for the reader's expectation of meaning change. The lower the error score, the more literal the expression. Metaphors are anomalous if no shared features are found (i.e. if the semantic representation vectors are orthogonal).

Nevertheless, this criterion may work for extreme cases of literal expressions (such as tautologies) or extreme anomalies, but not for metaphors. In the case of a tautology ('the apple is an apple'), indeed, the feature sets intersect completely, so that no different weights arise, and the error is probably about zero. Anomalies that share no feature at all are also easily detected, because the error will probably be about one. All cases in between - such as metaphors - are problematic. Certain expressions are literal with a strong metaphorical bias. In that case, error scores increase strongly, whereas readers still may judge them as literal (cf. Figure 6.2, Chapter 6). There are even metaphors with literal tendencies that may obtain smaller error scores than literal expressions with metaphoric tendencies, which is not expected by the MPC model.

What the MPC models misses is the determination of the nature of the error. The MPC model cannot tell what the optimal error value for metaphors is, because the features are not categorized as literal or figurative. Chapter 5 showed that this was crucial for the distinction between literal expressions and metaphors. Thus, two error scores could decide the expression type. Literal expressions would have a low error score for literal shared features, and a high one for (mixed) figurative shared features. Metaphors would have low scores on both dimensions, whereas anomalies would have two high scores.

Afterthoughts: Vivisection of the reader

Psycholinguistics and the theory of literature form a genuine cross-over. When literary theorists speculate on the effects of poetical devices or the identification of genres, they unavoidably assume a process that is responsible for that effect. Hence the need for psycholinguistics in the theory of literature. On the other hand, psycholinguistics cannot claim the validity of any language theory, as long as it cannot account for creative language processing. Hence the need for literary knowledge in psycholinguistics.

When hermeneuticians are entangled in the interpretation of poems, they convey nothing more than their psychological interaction with the text on an intuitive basis. When hermeneuticians want to escape from intuition as epistemological basis, they become empiricists.

An oft-repeated complaint from the hermeneuticians is that empiricists may succeed in avoiding sheer personal intuition as epistemological basis, but only at the cost of generalizing to an average interpretation of a work, while the really interesting part of literature is the unique reading experience. This is the opposition of 'mean versus variance'. The focus of the classic empiricist is on the mean value of a response, whereas the focus of the classic hermeneutician is on the variance of the mean, and particularly on the outliers (cf. Martindale 1991: 381-385). What is highly appreciated by the hermeneutician is the original insight of the brillant outlier, the new interpretation of an old text. What is highly appreciated by the empiricist is a smooth normal distribution. However, neither of them is right.

The validity of a general trend (the mean) in the reading population may only be derived from deviations from that mean trend (the variance). Reversely, a unique experience *cannot* be described if there is no implicit knowledge about the average. In other words, hermeneuticians do reckon with the mean. However, their assessment is unreliable, because it is not based on systematic analysis of random samples.

Tunnel vision on the mean is equally undesirable. Empiricists are predisposed to delete outliers from their data, supposedly 'belonging to another reading population', without ever returning to that 'other population' again. Variance is seen as a nuisance, and interactions that are insiduous in the experimental design are ignored to reduce it. However, interaction and variance are the supreme nuclei of human behavior.

Psycholinguists are not so much preoccupied with the covariates of the stimuli as they are with that of the subjects. This is already illustrated by the standard statistical packages, which provide the means of calculating the effects of covariates on the subjects level (e.g., for age or education), but not on the stimulus level (for word frequency or lexical ambiguity). In literature, the focus is the opposite, and this is what psycholinguistics can learn from literature. It is improper to match a stimulus set on, for example, word frequency in the language population and then to assume that it will be equal to word frequency in the subject group.

A more sensitive approach is needed, one which manages potential co-variates as a measure of the subject itself. In this vivisection of the reader, it is not assumed that a mean word frequency is applicable to every other subject from that population. What may be a high-frequent word in the population may be low-frequent for a specific subject, and vice versa. A word that is supposed to be ambiguous according to the dictionary may have such old meanings or infrequent uses that, de facto, it is an unambiguous word in the subject group. A psycholinguistic experiment, then, should start with pilots on subject-specific word frequencies, lexical ambiguity, similarity, and other relevant variables for the stimulus set, and investigate which stimuli have which particular problems for which subject. Afterwards, these measures can be used to correct the signal for that particular subject and stimuli. On the one hand, the suspicion that certain variables may interfere with a process is controlled less crudely than by correcting each subject with the same mean. On the other hand, effects may now be saved that - without the subject-specific correction - would have been regarded as too noisy.

This approach demands much more emphasis on the individual subject, who should be monitored during a longer period. It would also be advisable to keep different dependent variables within subjects (e.g., feature elicitation, scale value, RT, EEG), so that conclusions on their interrelationship can be drawn with more rigor. This would lead to the methodological requirement that the order of performing the tasks (including those performed in the pilots) are balanced across subjects and treated as a separate factor.

SUMMARY

Chapters 1, 2 and 3 introduce the pre-eminent schools of metaphor theory, thereby modeling the theories such that they become comparable. This is done by using a set theoretical approach, assuming that the *A*- and *B-term* in an expression activate one or more sets of semantic features, which should be seen as a sequence of associations.

The comparison model states that literal expressions have more shared features than metaphors, which in their turn have more shared features than anomalies. The size of the shared set was supposed to be checked in a one-stage process to determine the decision (Chapter 1).

The anomaly model (Chapter 2) states that one stage is needed for a literal expression, and two successive stages for a metaphor or anomaly. In stage 1, the *A*- and *B-term* would activate only literal features, which are then checked for the size of the shared set. If it turns out that the number of shared literal features is insufficient to make the expression literal, it would momentarily be perceived as an anomaly, which is accompanied by an electrocortical effect of semantic unexpectedness. An ensuing second stage would be needed to decide whether the expression is a metaphor. In stage 2, figurative features would be activated by the *A*- and *B-term*, and the size of the figurative shared set would determine whether the expression is actually a metaphor or yet an anomaly.

Chapter 3 introduces the interaction model, which also states that two stages are needed for a metaphor, and one for a literal expression. Literals would be processed as described for the comparison model, whereas metaphors would require a second stage in which relations are created for the *B-term* features. These relations would then be transferred to the *A-term* features, and if sufficient links are established, the expression would be a metaphor; otherwise, an anomaly. Moreover, interaction theory claims that adding context to the expressions would affect the process of creating relations.

To select a stimulus set that was natural for literary texts and yet fulfilled the demands of uniformity, Chapter 4 drew metaphors of the form 'the *A* is a *B*' ('the harbor is a mouth') from Dutch and Flemish poems. From these metaphors, literal ('the harbor is a place') and anomalous ('the harbor is a know-nothing') expressions were constructed. All *B-terms* met a host of constraints, so that expectations about spelling and rhyme were attenuated, numbers of syllables and letters were equal, word frequency differences and lexical ambiguity were avoided as far as possible, and syntactical form and function were standardized. In addition, these expressions were experimentally tested for category membership (is it literal, metaphoric or anomalous?) and for *A*- and *B-term* distribution (which is the focus, which is the referent?).

Chapter 5 tested the predictions on the elicited feature sets. Subjects produced features on the *A*- and *B-terms* in three conditions: On the isolated

terms in *single term*, on the terms in the isolated expression in *expression*, and on the terms in *context* (the original poem). Moreover, subjects interpreted the expressions in *expression* and *context*. Subsequently, they categorized the elicited semantic features as literal and/or figurative, and created relations among features and between features and terms. The predictions of the models as formalized in Chapters 1, 2 and 3 were compared with the results on feature production, and statistical inquiry determined which type of (shared) feature set had distinguished best among the expression types. It turned out that the shared sets defined by the comparison model could distinguish literals and metaphors from anomalies, but not from each other. Neither the number of created relations nor the number of shared relations could distinguish among expression types, which was not expected by the interaction model. The anomaly model did a little better in that literals and metaphors differed from anomalies for the size of the shared literal features, while literals and anomalies were different from metaphors for the size of the shared set of features that were literal for the one term and figurative for the other (mixed figurative, for short). *Context* and *expression* differed from *single term*, but not from each other. *Single term* evoked more features - yet smaller shared sets - whereas the opposite occurred for the other conditions.

Although certain shared sets were capable of determining the expression type, it was never the case that the interpretation of the expressions was identical to the shared sets. In other words, the interpretations contained many more features than merely the shared set.

Chapter 6 investigated the mental chronometry of metaphor processing. A two- and three-choice task were exploited to determine the speed with which literal expressions, metaphors and anomalies were processed in *expression* and *context*. Anomalies were processed fastest, followed by literal expressions and metaphors. The latter two could tie, with a slightly higher speed for literal expressions. There were no systematic differences between *expression* and *context*. It was concluded that these results can only be explained if the literal and figurative shared set are assumed to be established in parallel.

The main problem in this study was that severe interactions occurred with the order of presenting the expressions. The above was only valid for first presentations, whereas repetitions of the expressions erased all differences among expression types. Subjects probably recognized earlier judgements. Differences between the two- and the three-choice task also affected the processing. Subjects found it easier to decide between two choices than among three. Despite exhaustive practice, many errors occurred, due to poor discrimination among expression types rather than speed-accuracy problems.

Chapter 7 explored the electrocortical effects of reading a metaphor in *expression* and *context*. The anomaly model assumed that metaphor processing would be accompanied by a 'shock', due to the deviant meanings of the *A*- and *B-term*. This 'shock' was envisioned as the N400, an amplitude in the EEG that is supposed to reflect semantic mismatch. Indeed, the N400 occurred in the expected ordinal pattern, indicating that literal expressions evoked least N400-activity, metaphors intermediate, and anomalies the largest N400-activity, although the difference between the latter two was not statistically reliable. *Context* merely mitigated these differences.

However, the ordinal pattern for processing speed - anomalies < literal expressions ≤ metaphors - did not match the ordinal pattern for the N400-amplitude: Anomalies ≥ metaphors > literal expressions. It was deduced that N400 could not be part of the special metaphor processes. Instead, it might reflect an additional mismatch effect, based on category expectancy. Literal expressions match an instance with an appropriate category, whereas the categories for metaphors and anomalies are inappropriate. It might be, then, that the presence or absence of the N400 informs the reader about the literal status of an expression, which is confirmed by calculating the shared sets, thereafter.

In the conclusions, a new model of metaphor processing was proposed: The parallel two-stage anomaly race model. In this model, *A*- and *B-term* elicit a mix of features - not strictly serial - which are indexed as literal and/or figurative. Instances and categories are checked to see whether they match. If they do, the calculation of the literal shared set may be facilitated; if they do not, the N400 is evoked and the calculation of the (mixed) figurative shared set is facilitated. Subsequently, the literal shared set and the mixed figurative shared set are established in parallel by comparing each feature of the one term with each feature of the other. This is a race between the literal and mixed figurative shared set to satisfy the decision criteria. The expression type that meets 'end of file' first is the first to enter the response phase. In the response phase, expressions are first checked for a minimal size of the shared literal set, synchronously to a check for the minimal size of the shared mixed figurative features. If an expression remains below these minima, it is an anomaly. If the expression exceeds one of the minima, the next check asks whether the minimum for the shared mixed figurative set is met. If it is not, the expression is literal; if it is, the expression is a metaphor.

Ultimately, no model is definitive, and no experiment is the final test. Therefore, if the Tables and Figures turn out to be nothing more than rhetoric, then at least I hope that they may serve as a captatio for better researchers.

SAMENVATTING

Hoofdstuk 1, 2 en 3 van *De Metafoor en het Brein: Gedragsmatig en Psychofysiologisch Onderzoek naar Literaire Metafoorverwerking* behandelen de belangrijkste scholen in de metafoortheorie; de theorieën zo modellerend dat ze vergelijkbaar werden. De verzamelingstheoretische aanpak veronderstelde dat de *A*- en *B-term* van een expressie één of meer verzamelingen semantische kenmerken activeren, wat moet worden opgevat als één of meer reeksen associaties.

Het vergelijkingsmodel stelt dat letterlijke expressies meer gedeelde kenmerken hebben dan metaforen, die op hun beurt meer gedeelde kenmerken hebben dan anomalieën. De grootte van deze doorsnede wordt geacht gecontroleerd te worden in een één-stadiumproces ten einde tot een oordeel te komen (hoofdstuk 1).

Het anomaliemodel (hoofdstuk 2) stelt dat één stadium vereist is voor een letterlijke expressie en twee opeenvolgende stadia voor een metafoor en een anomalie. In stadium 1 zouden de *A*- en *B-term* alleen letterlijke kenmerken activeren, die vervolgens worden gecontroleerd op de grootte van de doorsnede. Als blijkt dat het aantal gedeelde letterlijke kenmerken onvoldoende is om de expressie letterlijk te maken, zou deze even opgevat worden als anomalie, begeleid door een electrocorticaal effect van semantische onverwachtheid. Een daaropvolgend tweede stadium zou nodig zijn om te beslissen of de expressie een metafoor is. In stadium 2 zouden figuurlijke kenmerken geactiveerd worden door de *A*- en *B-term* en de grootte van de figuurlijke doorsnede zou bepalen of de expressie daadwerkelijk een metafoor is of toch een anomalie.

Hoofdstuk 3 introduceert het interactiemodel, dat eveneens stelt dat twee stadia benodigd zijn voor een metafoor en één voor een letterlijke expressie. Letterlijken zouden verwerkt worden zoals beschreven in het vergelijkingsmodel, terwijl metaforen een tweede stadium zouden vergen waarin relaties gecreëerd worden voor de kenmerken van de *B-term*. Deze relaties zouden dan overgeheveld worden naar de kenmerken van de *A-term* en als er genoeg verbanden zijn gelegd, zou de expressie een metafoor zijn; anders een anomalie. Bovendien beweert de interactietheorie dat het toevoegen van context aan de expressies het proces van relatiecreatie zou beïnvloeden.

Om een stimulusverzameling te selecteren die natuurlijk was voor literaire teksten en toch voldeed aan de eisen van uniformiteit, onttrok hoofdstuk 4 metaforen van de vorm 'de *A* is een *B*' ('de haven is een mond') aan Nederlandse en Vlaamse gedichten. Uit deze metaforen werden letterlijke ('de haven is een plek') en anomale ('de haven is een snars') expressies geconstrueerd. Alle *B-termen* voldeden aan een reeks eisen, zodanig dat verwachtingen over spelling en rijm werden gematigd, het aantal lettergrepen en letters gelijk was, verschillen in woordfrequentie en lexicale ambiguïteit zoveel mogelijk werden vermeden en grammaticale vorm en functie werden gestandaardiseerd. Daarenboven werden deze expressies experimenteel getoetst op hun categoriebepaaldheid (letterlijk, metaforisch of anomaal?) en op *A*- en *B-term*-distributie (wat is de focus, wat is de referent?).

Hoofdstuk 5 testte de voorspellingen over de opgeroepen kenmerkverzamelingen. Proefpersonen produceerden kenmerken voor de *A*- en *B-termen* in drie condities: voor de geïsoleerde termen in *losse term*, voor de termen in de geïsoleerde expressie in *expressie* en voor de termen in *context* (het oorspronkelijke gedicht). Bovendien interpreteerden de proefpersonen de expressies in *expressie* en *context*. Vervolgens categoriseerden zij de gegenereerde semantische kenmerken als letterlijk en/of figuurlijk en creëerden relaties tussen kenmerken en tussen kenmerken en termen. De voorspellingen van de modellen zoals geformaliseerd in de hoofdstukken 1, 2 en 3 werden vergeleken met de resultaten van de kenmerkproductie en statistische analyse bepaalde welk type verzameling het best de expressietypen kon onderscheiden. Het bleek dat de doorsneden gedefinieerd door het vergelijkingsmodel, letterlijken en metaforen konden onderscheiden van anomalieën, maar niet van elkaar. Het aantal gecreëerde relaties noch het aantal gedeelde relaties konden expressietypen onderscheiden, wat niet verwacht werd door het interactiemodel. Het anomaliemodel deed het iets beter, omdat letterlijken en metaforen verschilden van anomalieën in de grootte van de letterlijke doorsnede, terwijl letterlijken en anomalieën verschillend waren van metaforen in het aantal gedeelde kenmerken dat letterlijk was voor de ene term en figuurlijk voor de andere (afgekort als 'gemengd figuurlijk'). *Context* en *expressie* verschilden van *losse term*, maar niet van elkaar. *Losse term* wekte meer kenmerken op, maar de doorsneden waren kleiner, terwijl het omgekeerde gebeurde voor de andere condities.

Hoewel bepaalde doorsneden in staat waren het expressietype te determineren, was het nooit zo dat de interpretatie van de expressies identiek was aan de doorsneden. Met andere woorden, de interpretaties bevatten veel meer kenmerken dan alleen de doorsnede.

Hoofdstuk 6 onderzocht de mentale chronometrie van metafoorverwerking. Een twee- en driekeuzentaak werden benut om de snelheid te bepalen waarmee letterlijke expressies, metaforen en anomalieën verwerkt werden in *expressie* en *context*. Anomalieën werden het snelst verwerkt, gevolgd door letterlijke expressies en metaforen. De laatste twee konden gelijk eindigen, met soms een iets hogere snelheid voor letterlijke expressies. Er waren geen systematische verschillen tussen *expressie* en *context*. Er werd geconcludeerd dat deze resultaten alleen te verklaren zijn wanneer de letterlijke en figuurlijke doorsnede verondersteld worden parallel tot stand te komen.

Het grootste probleem in deze studie was dat ernstige interacties optraden met de volgorde van presentatie van de expressies. Bovenstaande was alleen geldig voor eerste presentaties, terwijl herhalingen van de expressies alle verschillen tussen de expressietypen uitwisten. Waarschijnlijk herkenden de proefpersonen hun eerdere beslissingen. Verschillen tussen de twee- en driekeuzentaak beïnvloedden eveneens de verwerking. Proefpersonen vonden het eenvoudiger te beslissen tussen twee keuzen dan tussen drie. Ondanks langdurige oefening traden vele vergissingen op, eerder door de slechte discrimineerbaarheid van de expressietypen dan door snelheid-accuraatheids-problemen.

Hoofdstuk 7 exploreerde de electrocorticale effecten van het lezen van een metafoor in *expressie* en *context*. Het anomaliemodel veronderstelde dat metafoorverwerking zou worden begeleid door een 'schok', veroorzaakt door de afwijkende betekenissen van de *A*- en *B-term*. Deze 'schok' werd voorgesteld als de N400, een amplitude in het EEG die verondersteld wordt semantische ongelijkheid te representeren. Inderdaad trad de N400 op volgens het verwachte ordinale patroon, uitwijzend dat letterlijke expressies de minste N400-activiteit uitlokten, metaforen gematigde en anomalieën de grootste N400-activiteit, hoewel het verschil tussen de laatste twee statistisch niet betrouwbaar was. *Context* verkleinde slechts deze verschillen.

Echter, het ordinale patroon voor de verwerkingssnelheid (anomalieën $<$ letterlijke expressies $\leq$ metaforen) kwam niet overeen met het ordinale patroon voor de N400-amplitude: anomalieën $\geq$ metaforen $>$ letterlijke expressies. Gededuceerd werd dat N400 geen onderdeel van het speciale metafoorproces kon zijn, maar dat het een toegevoegd effect van ongelijkheid kan zijn, gebaseerd op een categorieverwachting. Letterlijke expressies combineren een exemplaar met een toepasselijke categorie, terwijl de categorieën van metaforen en anomalieën niet toepasselijk zijn. Het kan aldus zijn dat de aan- of afwezigheid van de N400 de lezer informeert over de letterlijke status van een expressie, die bevestigd wordt door daarna de doorsnede te berekenen.

In de conclusies werd een nieuw model voor metafoorverwerking voorgesteld: het parallelle twee-stadia anomalie wedloopmodel. In dit model wekken *A*- and *B-term* een mengsel van kenmerken op, niet strikt serieel, die geïndiceerd worden als letterlijk en/of figuurlijk. Gecontroleerd wordt of exemplaren en categorieën overeenstemmen. Als dat zo is, kan de berekening van de letterlijke doorsnede gefaciliteerd worden; wanneer niet, wordt de N400 opgewekt en de berekening van de (gemengd) figuurlijke doorsnede gefaciliteerd. Vervolgens komen de letterlijke en gemengd figuurlijke doorsnede parallel tot stand door elk kenmerk van de ene term te vergelijken met elk kenmerk van de andere. Dit is een wedloop tussen de letterlijke en gemengd figuurlijke doorsnede om te voldoen aan de decisiecriteria. Het expressietype dat het eerst 'einde bestand' bereikt gaat het eerste de responsfase in. In de responsfase worden expressies eerst gecontroleerd op een minimumomvang voor de letterlijke doorsnede synchroon aan een controle op de minimumomvang van de gemengd figuurlijke doorsnede. Als een expressie beneden deze minima blijft, is het een anomalie. Als een expressie één van deze minima overstijgt, vraagt de volgende controle of het omvangsminimum voor de gemengd figuurlijke doorsnede wordt gehaald. Als dat niet het geval is, is de expressie letterlijk; wanneer wel, dan is de expressie een metafoor.

Uiteindelijk is geen model definitief en geen experiment de doorslaggevende test. Mochten derhalve de tabellen en figuren niet meer dan retorica blijken te zijn, dan hoop ik tenminste dat ze als een captatio mogen dienen voor betere onderzoekers.

REFERENCES

Abernathy, R. (1967). Rhymes, non-rhymes, and antirhyme. In: *To honor Roman Jakobson, essays on the occasion of his seventieth birthday*, I. Mouton, The Haque, Paris.

Adrian, E.D., & Matthews, B. (1934). The Berger rhythm: Potential changes from the occipital lobes in man. *Brain*, *57*, 355-387.

Anderson, C.C. (1964). The psychology of the metaphor. *The Journal of Genetic Psychology*, *105*, 53-73.

Baayen, R.H., Piepenbrock, R., & Rijn, H., van (1993). *The CELEX Lexical Database*. Linguistic Data Consortium, Univ. of Pennsylvania, Philadelphia, Pa.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, *3*, 340-357.

Barthes, R. (1970). *Writing Degree Zero and Elements of Semiology*. Beacon Press, Boston.

Beardsley, M.C. (1982). The metaphorical twist. In: *The Aesthetic Point of View*. M.J. Wreen & D.M. Callen (Eds.). Cornell University Press, Ithaca, London.

Becker, C.A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 252-259.

Becker, C.A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition*, *8*, 493-512.

Benson, D.F. (1981). Alexia and the neuroanatomical basis of reading. In: *Neuropsychological and Cognitive Processes in Reading*. F.J. Pirozzolo & M.C. Wittrock (Eds.). Academic Press, New York.

Bentin, S., McCarthy, G., & Wood, C.C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343-355.

Berger, H. (1929). Über das Electrienenkephalogramm des Menschen. *Archiv für Psychiatrie*, *87*, 527-570.

Beringer, J. (1992). *Entwerf einer Anwendersprache zur Steuerung psychologischer Reaktionszeitexperimente*. Peter Lang, Frankfurt am Main.

Berlyne, D.E. (1960). *Conflict, Curiosity and Arousal*. McGraw-Hill, New York.

Berlyne, D.E. (1971). *Aesthetics and Psychobiology*. Meredith, New York.

Black, D.L. (1990). *Logic and Aristotle's Rhetoric and Poetics in medieval Arabic philosophy. Islamic philosophy and theology. Texts and studies*, *7*. H. Daiber (Ed.). E.J. Brill, Leiden, New York, Kobenhavn, Köln.

Black, M. (1962). *Models and Metaphors*. Cornell University Press, Ithaca, New York.

Black, M. (1979). More about metaphor. In: *Metaphor and Thought*. A. Ortony (Ed.). Cambridge University Press, Cambridge.

Boddy, J., & Weinberg, H. (1981). Brain potentials, perceptual mechanisms and semantic categorisation. *Biological Psychology*, *12*, 43-61.

Boogaart, P.C., uit den (Ed.) (1975). *Woordfrequentielijst in geschreven en gesproken Nederlands*. Oosthoek, Scheltema & Holkema, Utrecht.

Breton, A. (1955). *Les Vases Communicants*. Gallimard, Paris.

Burke, K. (1945). *A Grammar of Motives*. Prentice Hall Inc., Englewood Cliffs, New Jersey, pp. 503-504.

Burnage, G. (1990). *CELEX: A Guide for Users*. CELEX, Nijmegen, Holland.

Campbell, K. & Bell, I. (1992). *InstEP systems*. InstEP corp. Ottawa, Canada.

Changeux, J.P. (1985). *Neuronal Man. The Biology of Mind*. Oxford University Press, New York.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335-359.

Clark, H.H. & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, *14*, 56-72.

Culler, J. (1981). *The Pursuit of Signs: Semiotics, Literature, Deconstruction*. Routledge & Kegan Paul, London, Henley.

Chwilla, D.J., Brown, C.M. & Hagoort, P. (1995). The N400 as a function of the level of processsing. *Psychophysiology*, *32*, 274-285.

Dale, van (1984). *Groot woordenboek der Nederlandse Taal*. G. Geerts, H. Heestermans & C. Kruyskamp (Eds.). Van Dale Lexicografie, Utrecht, Antwerpen.

Danesi, M. (1989). The neurological coordinates of metaphor. *Communication & Cognition*, *22*, *1*, 73-86.

Donders, F.C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologische Laboratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II, 92-120. Translation by W.G. Koster, *Acta Psychologica*, *30*, Attention and performance, 1969, 421-431.

Eco, U. (1972). *Einführung in die Semiotik*. W. Fink Verlag, München.

Eco, U. (1979). *The Role of the Reader*. Indiana University Press, Bloomington.

Estill, R.B., & Kemper, S. (1982). Interpreting idioms. *Journal of Psycholinguistic Research*, *11*, *6*. 559-568.

Fischler, I., & Bloom, P.A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 1-20.

Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E., & Perry, N.W. jr. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*, *4*, 400-409.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In: *Sentence processing*. W.E. Cooper & E.C.T. Walker (Eds.). Erlbaum, Hillsdale, NJ.

Foucault, M. (1988). *Dit is geen pijp*. Met reproducties en brieven van René Magritte. C. Heering-Moorman (Ed.). Aramith Uitgevers, Amsterdam. Translation from *Ceci n'est pas une pipe*. Fata Morgana, St. Clement, 1970.

Flesh, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*, *3*, 221-233.

Garner, W.R. (1962). *Uncertainty and Structure as Psychological Concepts*. Wiley, New York.

Genette, G. (1970). La rhétorique restreinte. *Figures*, *III*. Éditions du Seuil, Paris.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*, 1, 49-59.

Gentner, D. (1989). The mechanisms of analogical learning. In: *Similarity and Analogical Reasoning*. S. Vosniadou & A. Ortony (Eds.). Cambridge University Press, Cambridge, New York, New Rochelle, Melbourne, Sydney, 199-233.

Gerrig R.J. (1989_a). Empirical constraints on computational theories of metaphor: Comments on Indurkhya. *Cognitive Science*, *13*, 235-241.

Gerrig, R.J. (1989_b). The time course of sense creation. *Memory and Cognition*, *17*, *2*, 194-207.

Gerrig, R.J., & Healy, A.F. (1983). Dual processing in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 667-675.

Gibbs, R.W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition*, *8*, *2*, 149-156.

Gibbs, R.W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge University Press. Cambridge, New York, Melbourne.

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior*, *22*, 577-590.

Gleitman, L.R., Gleitman, H., Miller, C. & Ostrin, R. (1996). Similar, and similar concepts. *Cognition*, *58*, 321-376.

Glucksberg, S., Gildea, P., & Bookin, H.B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, *21*, 85-98.

Glucksberg, S., Kreuz, R.J., & Rho, S.H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 323-335.

Glucksberg, S., Brown, M., & McGlone, M.S. (1993). Conceptual metaphors are not automatically accessed during idiom comprehension. *Memory & Cognition*, *21*, *5*, 711-719.

Goldberg, E., & Costa, L.D. (1981). Hemispheric differences in the acquisition of descriptive systems. *Brain and Language*, *14*, 144-173.

Goldstone, R.L., Medin, D.L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgements. *Cognitive Psychology*, *23*, 222-262.

Greimas, A.J. (1966). *Sémantique Structurale, Recherche de Méthode*. Larousse, Paris.

Halliwell, S. (1987). *The Poetics of Aristotle, Translation and Commentary*. The University of North Carolina Press. Chapel Hill, USA.

Harris, R. (1976). Comprehension of metaphor: A test of a two-stage processing model. *Bulletin of the Psychonomic Society*, *8*, 312-324.

Hasher, L., & Zacks, R.T. (1979). Automatic and effortful processing in memory. *Journal of Experimental Psychology: General*, *108*, 356-388.

Hayward, M. (1996). Analysis of a corpus of poetry by a connectionist model of poetic meter. *Poetics*, *24*, 1-11.

Henle, P. (1966). *Language, Thought and Culture*. Ann Arbor, University of Michigan Press, USA.

Hillyard, S., & Kutas, M. (1983). Electrophysiology of Cognitive Processing. *Annual Review of Psychology*, *34*, 33-61.

Hoffman, R.R., & Kemper, S. (1987). What could reaction-time studies be telling us about metaphor comprehension? *Metaphor and Symbolic Activity*, *2*, 149-186.

Hoorn, J.F. (1996). Psychophysiology and literary processing: ERPs to semantic and phonological deviations in reading small verses. In: *Empirical Approaches to Literature and Aesthetics*. R.J. Kreuz & M.S. MacNealy (Eds.). Ablex Publishing Corporation, Norwood, New Jersey.

Hoorn, J.F. (in press). Modeling metaphor theory, behavioral underscoring, and psychophysiological backgrounds. In: *Cognitive Aspects of Metaphor: Structure, Comprehension, and Use*. Y. Shen & A. Kasher (Eds.).

Hoorn, J.F., & Veldman, O.C.J. (in preparation). Perceiving shared features in two word-and image paintings of René Magritte.

Hopkins, G.M. (1966). *The Journals and Papers of Gerard Manley Hopkins*. H. House & G. Storey (Eds.). Oxford University Press, London, New York, Toronto.

Huang, K. (1992). *Quarks, Leptons & Gauge Fields*. World Scientific, Singapore, New Jersey, London, Hong Kong.

Hynd, G.W. (1989). Learning disabilities and neuropsychological correlates: Relationship to Neurobiological Theory. In: *Learning Disabilities*, *2*. D.J. Bakker & H. van der Vlugt (Eds.). Zwets & Zeitlinger, Lisse, The Netherlands.

Ibsch, E. (1991). Opening Address: An oral introduction. In: *Empirical Studies of Literature: Proceedings of the Second IGEL-Conference, Amsterdam 1989*. E. Ibsch, D. Schram & G. Steen (Eds.). Rodopi, Amsterdam, Atlanta.

Inhoff, A.W., Lima, S.D., & Carroll, P.J. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition*, *12*, 558-567.

Jakobson, R. (1960). Linguistics and poetics. In: *Style in Language*. T. Sebeok (Ed.). MIT Press, Cambridge, Mass.

Jakobson, R. (1961). Linguistics and communication theory. In: *Proceedings of Symposia in Applied Mathematics*, *12*. American mathematical society, Providence, Rhode Island.

Jakobson, R., Fant, C.G.M., & Halle, M. (1963). *Preliminaries to Speech Analysis. The Distinctive Features and their Correlates*. MIT Press, Cambridge, Massachusetts.

Jakobson, R. (1971). *Selected Writings, II, Word and Language*. Mouton, The Haque, Paris.

Jakobson, R. (1973). *Questions de Poetique*. Seuil, Paris.

Jakobson, R. (1981). *Selected Writings, III, Poetry of Grammar and Grammer of Poetry*. Mouton, The Hague, Paris, New York.

James, C.T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 130-136.

Janus, R.A., & Bever, T.G. (1985). Processing of metaphoric language: An investigation of the three-stage model of metaphor comprehension. *Journal of Psycholinguistic Research*, *14*, *5*, 473-487.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography & Clinical Neuropsychology*, *10*, 371-375.

Jastrzembski, J. (1981). Multiple meanings, number of related meanings, frequency of occurence, and the lexicon. *Cognitive Psychology*, *13*, 278-305.

Johnson, M.G., & Malgady R.G. (1979). Some cognitive aspects of figurative language: Association and metaphor. *Journal of Psycholinguistic Research*, *8*, 249-266.

Katz, A.N. (1982). Metaphoric relationships: The role of feature saliency. *Journal of Psycholinguistic Research*, *11*, *4*, 283-296.

Kintsch, W. (1974). *The Representation of Meaning in Memory*. Lawrence Erlbaum Ass. Publishers, Hillsdale, New Jersey.

Kintsch, W., & Mross, E.F. (1985). Context effects in word identification. *Journal of Memory and Language*, *24*, 336-349.

Klare, G.R. (1963). *The Measurement of Readability*. The Iowa State University Press, Ames, Iowa, USA.

Kok, A. (1988). *De Psychologische Betekenis van Hersenpotentialen*. Inaugurele rede UvA. Sneldruk Enschede.

Kramer, A., & Donchin, E. (1987). Brain potentials as indices of orthographic and phonological interaction during word matching. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 76-86.

Kreitler, H. & Kreitler, S. (1972). *Psychology of the Arts*. Duke University Press, Durham, N.C.

Krumhansl, C. (1978). Concerning the applicability of geometrical models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *85*, 445-463.

Kutas, M., & Hillyard, S. (1980_a). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Kutas, M., & Hillyard, S. (1980_b). Reading between the lines: Event-related potentials during natural sentence processing. *Brain and Language*, *11*, 354-373.

Kutas, M., & Hillyard, S. (1981). ERPs to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*, 99-116.

Kutas, M., & Hillyard, S. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologica*, *20*, 579-590.

Kutas, M., & Hillyard, S. (1983). ERPs to grammatical errors and semantic anomalies. *Memory and Cognition*, *11*, 5, 539-550.

Kutas, M., Lindamood, T.E., & Hillyard, S.A. (1984). Word expectancy and ERP during sentence processing. In: *Preparatory States and Processes*. S. Kornblum & J. Requin (Eds.). LEA, Hillsdale, New Yersey, London, 217-237.

Levin, S.R. (1977). *The Semantics of Metaphor*. J. Hopkins University Press, Baltimore, London.

Lotman, Y.M. (1972). *Die Struktur Literarischer Texte*. Wilhelm Fink Verlag, München.

Lotman, Y.M. (1976). *Analysis of the Poetic Text*. Ardis, Ann Arbor.

Loveland, W., & Seaborg, G. (1991). The search for the missing elements. *New Scientist*, *31*, 29-32.

Malgady, N., & Johnson, M. (1980). Measurement of figurative language: Semantic feature models of comprehension and appreciation. In: *Cognition and Figurative Language*. R. Honeck & R. Hoffman (Eds.). Hillsdale, New Yersey, 239-258.

Martin, J.H. (1992). Computer understanding of conventional metaphoric language. *Cognitive Science*, *16*, 233-270.

Martindale, C. (1991). Chaos theory, strange attractors, and the laws of literary history. In: *Empirical Studies of Literature: Proceedings of the Second IGEL-Conference, Amsterdam 1989*. E. Ibsch, D. Schram & G. Steen (Eds.). Rodopi, Amsterdam, Atlanta.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287-330.

Meyer, D.E., Schvaneveldt, R.W., & Ruddy, M.G. (1975). Loci on contextual effects on visual word recognition. In: *Attention and Performance*, *V*. P.M.A. Rabbitt & S. Dornii (Eds.). Academic Press, NY, London.

Meyer, D.E., Osman, A.M., Irwin, D.E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, *26*, 3-67.

Miall, D.S. (1995). Anticipation and feeling in literary response: A neuropsychological perspective. *Poetics*, *23*, 275-298.

Miller, G.A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology 6*, 169-191.

Miller, G.A., & Johnson-Laird, P.N (1976). *Language and Perception*. Harvard Univ. Press, Cambridge, Massachusetts.

Millward, R.B., Rice, G., & Corbett, A. (1975). Category production measures and verification times. In: *Studies in Long-Term Memory*. R.A. Kennedy & R.W. Wilkes (Eds.). Wiley, New York, 219-252.

Mooij, J.J.A. (1976). *A Study of Metaphor: On the Nature of Metaphorical Expressions, with Special Reference to their Reference*. North Holland, Amsterdam.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165-178.

Mukařovský, J. (1964). Standard language and poetic language. In: *A Prague School Reader on Aesthetics, Literary Structure, and Style*. P.I. Garvin (Ed.). Washington, Georgetown University Press.

Nagy, M.E. & Rugg, M.D. (1989). Modulation of event-related potentials by word repetition: The effect of inter-item lag. *Psychophysiology*, *26*, *4*, 431-436.

Nambu, Y. (1985). *Quarks, Frontiers in Elementary Particle Physics*. World Scientific, Philadelphia, Singapore.

Nigram, A., Hoffman, J.E., & Simons, R.F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, *4*, 15-22.

Nunez, P. (1981). *Electric Fields of the Brain*. Oxford University Press, New York, Oxford.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, *97*, *2*, 316-333.

Onifer, W., & Swinney, D. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition*, *9*, 225-236.

Ortony, A., Schallert, D.L., Reynolds, R.E., & Antos, S.J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehesion. *Journal of Verbal Learning and Verbal Behaviour*, *17*, 465-477.

Ortony, A. (1979). The role of similarity in similes and metaphors. In: *Metaphor and Thought*. A. Ortony (Ed.). Cambridge University Press, New York.

Ortony, A., Vondruska, R., Foss, M., & Jones, L. (1985). Salience, similes, and the asymmetry of similarity. *Journal of memory and Language*, *24*, 569-594.

Osgood, C.E., Suci, G.J. & Tannenbaum, P.H. (1957). *The Measurement of Meaning*. Urbana, University of Urbana Press.

Paivio, A., & Clark, J.M. (1986). The role of topic and vehicle imagery in metaphor comprehension. *Communication and Cognition*, *19*, *3, 4*, 367-388.

Petten, C. van, & Kutas, M. (1987). Ambiguous words in context: An event-related potential analyses of the time course of meaning activation. *Journal of Memory and Language*, *26*, 188-208.

Polich, J., McCarthy, G., Wang, W., & Donchin, E. (1983). When words collide: Orthographic and phonological interference during word processing. *Biological Psychology*, *16*, 155-180.

Polich, J. (1985). Semantic categorization and event-related potentials. *Brain & Language*, *26*, *2*, 304-321.

Posner, M.I., & Snyder, C.R. R. (1975). Attention and cognitive control. In: *Information Processing and Cognition: The Loyola Symposium*. R.L. Solso (Ed.). Erlbaum, Hillsdale, NJ.

Posner, M.I. (1978). *Chronometric Explorations of Mind*. Hillsdale, NJ: Erlbaum.

Postman, L., & Conger, B. (1954). Verbal habits and the visual recognition of words. *Science*, *119*, 671-673.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1980). SAM: A theory of probabilistic search of associative memory. In: *The Psychology of Learning and Motivation: Advances in Research and Theory*, *14*. G.H. Bower (Ed.). Academic Press, New York.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.

Raaijmakers, J.G.W. (1984). *Psychologie van het Geheugen*. Van Loghem Slaterus, Deventer.

Ratcliff, R. (1988). Continuous versus discrete information processing. Modeling accumulation of partial information. *Psychological Review*, *95*, 238-254.

Rhys Roberts, W. (1966). *The Works of Aristotle*, *11*, Oxford, UK.

Richards, I.A. (1965). *The Philosophy of Rhetoric*. Oxford University Press, London, Oxford, New York.

Ricœur, P. (1974). Metaphor and the main problem of hermeneutics. *New Literary History*, *VI*, *1*, 95-110.

Riffaterre, M. (1980). *Semiotics of Poetry*. Methuen, London.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In: *Cognitive Development and the Acquisition of Language*. T.E. Moore (Ed.). Academic Press, New York.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532-547.

Rosch, E., & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-603.

Rosch, E., Mervis, C.B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

Roth, N. & Boddy, J. (1989). Event-related potentials and the recognition of subliminally exposed words after repeated presentation. *Journal of Psychophysiology*, *3*, 281-289.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, *53*, 94-101.

Rubenstein, H., Garfield, L., & Millikan, J.A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*, 487-494.

Rubenstein, H., Lewis, S.S., & Rubenstein, M.A. (1971). Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior*, *10*, 57-62.

Rugg, M.D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language*, *23*, 225-240.

Rugg, M.D. (1985). The effects of semantic priming and word repetition on the modulation of event-related potentials by word repetition. *Psychophysiology*, *25*, 55-63.

Rugg, M.D., Kok, A., Barrett, S.E., & Fischler, I. (1985). ERPs associated with language and hemispheric specialization. Supplement of Electroencephalography and Clinical Neurophysiology: *Proceedings of the Seventh International Conference on Event Related Potentials of the Brain*.

Rugg, M.D. (1987). Dissociation of semantic priming, word and non-word repetition effects by event-related potentials. *The Quarterly Journal of Experimental Psychology*, *39A*, 123-148.

Rugg, M.D., & Barrett, S.E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language*, *32*, 336-361.

Rugg, M.D., & Nagy, M.E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory & Cognition*, *15*, *6*, 473-481.

Rugg, M.D., Furda, J. & Lorist, M. (1988). The effects of task on the modulation of event-related potentials by word repetition. *Psychophysiology*, *25*, *1*, 55-63.

Rumelhart D.E., & Ortony, A. (1977). The representation of knowledge in memory. In: *Schooling and the Acquisition of Knowledge*. R.C. Anderson, R.J. Spiro & W.E. Montaque (Eds.). Erlbaum, Hillsdale, NJ.

Sanders, A.F. (1980). Stage analysis of reaction processes. In: *Tutorials in Motor Behavior*. G.E. Stelmach & J. Requin (Eds.). North Holland, Amsterdam.

Sanders, A.F. (in press). *Elements of Human Performance*. Erlbaum, Hillsdale, NJ.

Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K., & Lindsley, D.B. (1980). Electrocortical sins of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology*, *17*, *6*, 568-576.

Scarborough, D.L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory and Cognition*, *7*, 3-12.

Schvaneveldt, R.W., Meyer, D.E., & Becker, C.A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 243-256.

Seidenberg, M.S., & Tanenhaus, M.K. (1979). Orthographic effects of rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 546-554.

Seidenberg, M., Tanenhaus, M., Leiman, J., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*, 489-537.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing I. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127-190.

Simpson, G.B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*, 120-136.

Simpson, G.B., & Burgess, C. (1982). Processing lexical ambiguity: A chronometric analysis. *Paper Psychonomic Society*, Minneapolis, Minnesota.

Simpson, G.B. (1984). Lexical ambiguity and its role in models of word recognition, *Psychological Bulletin*, *96*, 316-340.

Skinner, B.F. (1957). *Verbal behaviour*. Methuan & Co. Ltd., London.

Šklovskij, V. (1965). Art as technique. In: *Russian Formalist Criticism: Four Essays*. L.T. Lemon & M. Reis (Eds.). Nebraska University Press, Lincoln.

Solomon, R.L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experiental Psychology*, *43*, 195-201.

Sonesson, G. (1996). An essay concerning images: From rhetoric to semiotics by way of ecological physics. *Semiotica*, *109*, *1*, *2*, 41-140.

Steen, G. (1994). *Understanding Metaphor in Literature: An Empirical Approach*. Longman, London, New York.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, *30*, Attention and performance 2, 276-315.

Stevens, S.S., & Stone, G. (1947). Psychological writing, easy and hard. *American Psychologist*, *2*, 230-235, 523-525.

Striedter, J. (1989). *Literary Structure, Evolution, and Value: Russian Formalism and Czech Structuralism Reconsidered*. Harvard University Press, Cambridge.

Stuss, D.T., Sarazin, F.F., Leech, E.E., & Picton, T.W. (1983). Event-related potentials during naming and mental rotation. *Electroencephalography and Clinical Neuropsychology*, *56*, 133-146.

Stuss, D.T., Picton, T.W., & Cherri, A.M. (1986). Searching for the names of pictures: An event-related potential study. *Psychophysiology*, *23*, *2*, 215-223.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 645-660.

Tanenhaus, M., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 427-440.

Tanenhaus, M.K., Flanigan, H., & Seidenberg, M.S. (1980). Orthographic and phonological code activation in auditory and visual word recognition. *Memory and Cognition*, *8*, 513-520.

Thomas, M.S.C. & Mareschal, D. (in press). A connectionist model of metaphor pattern completion. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, 696-701.

Thurstone, L.L. (1927). A law of comparative judgement. *Psychological Review*, *34*, 273-386.

Till, R.E., Mross, E.F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory and Cognition*, *16*, *4*, 283-298.

Tourangeau, R., & Sternberg, R. (1982). Understanding and appreciating metaphors. *Cognition*, *11*, 203-244.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, *84*, *4*, 327-352.

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, *24*, *4*, 479-486.

Verbrugge, R., & McCarrell N. (1977). Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, *9*, 494-533.

Wellek, R., & Warren, A. (1955). *Theory of literature*. J. Cape, London.

Whaley, C.P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behaviour*, *17*, 143-154.

Winner, E., & Gardner, H. (1977). The comprehension of metaphor in brain-damaged patients. *Brain*, *100*, 717-729.

Woestenburg, J.C., Verbaten, M.N., Hees, van H.H. & Slangen, J.L. (1983). Single trial ERP estimation in the frequency domain using Orthogonal Polynomial Trend Analysis (OPTA): Estimation of individual habituation. *Biological Psychology*, *17*, 173-191.

Woestenburg, J.C., Das-Smaal, E.A., Brand, E., & Kramer, S. (1992). Learning during visual search in children with attentional and learning problems: A trial-to-trial evaluation of RT and ERP measure. *Journal of Psychophysiology*, *6*, 204-224.

Woestenburg, J.C. (1994). *Statistical Timeseries Package for Biological Signals (STPBS)*. *Talo b.v., Bussum, The Netherlands.

Wolff, R. (1977). *Strukturalismus und Assoziationspsychology. Empirisch-pragmatische Literaturwissenschaft im Experiment: Baudelaires "Les Chats"*. Max Niemeyer Verlag. Tübingen.

Zwaan, R.A. (1991). The literary situational-action model and the reading process. In: *Empirical Studies of Literature: Proceedings of the Second IGEL-Conference, Amsterdam 1989*. E. Ibsch, D. Schram & G. Steen (Eds.). Rodopi, Amsterdam, Atlanta.

Zwaan, R.A. (1993). *Aspects of Literary Comprehension: A Cognitive Approach*. Benjamins, Amsterdam, Philadelphia.

AUTHOR INDEX

GLOSSARY - INDEX OF KEY WORDS

stimulus - 135, 142-145, 159

Feature set: A number of associations summed up in response to a stimulus - 10, 18-20, 22, 25, 35, 37, 39, 42, 48-50, 52, 54, 58, 59, 61, 82, 85, 87, 95, 96, 102, 120, 121, 126-128, 130, 133-138, 140-142, 145, 147, 148, 150, 153-155, 158, 167, 168, 171, 174, 175, 187, 190, 192-195, 203, 209, 212, 215, 222, 228, 302, 307, 316, 323, 329

Figurative features: Those features in the set that are judged as imaginative, metaphoric or symbolic - 30, 35, 39, 42, 43, 147, 155, 168, 169, 200

Figurativeness: The nonliteral resemblance between stimuli, supposedly based on a measure of their figurative and nonfigurative features - 120, 135, 141, 142, 187-189, 206, 207, 225

Foregrounding: Textual elements that are deviant from expectation, evoking surprise effects - 72, 232, 233, 269, 272

Formal equality: The exact identity of two letter strings - 146, 147, 205

Grand mean: The mean of individual means across subjects - 153-166, 168, 174, 175, 179, 183, 187, 188, 191, 193-195, 197, 215, 218, 225, 226, 247-250, 252-256, 263, 266, 268, 285, 288, 290, 292, 296, 299, 301, 309

Home key: Reaction time device to make comparable the multidirectional decisions of one finger. Since the home key must be released before a decision key is pressed, the spatial or motor preparation differences are attenuated that may be predicated by the position of the decision keys - 246, 247, 283, 284

Homonymy: Cf. lexical ambiguity - 73, 78, 82, 101

Idiom: Metaphor that is incorporated in the standard language, and has lost its literal meaning - 236, 238, 241, 242

Independent variable: A preferably stable phenomenon that is manipulated to study its effect on the dependent variable. Cf. **factor**.

Instance: Member of a category of objects - 9, 14, 19, 25, 31, 36, 37, 51, 52, 61, 65, 74-76, 78, 81-86, 88, 96, 97, 99, 100, 102, 103, 119, 122, 128, 132, 134, 140, 141, 189, 192, 196-199, 202, 205, 232, 236, 271-273, 275, 277, 278, 301-303, 306, 316, 322, 325, 330

Instance-category verification: Experimental task in which subjects judge whether instances and categories match - 81, 82, 85, 102, 103, 128, 196

Instance dominance: Number of times an instance is mentioned in response to the category - 81, 82, 85, 102, 103, 128, 196

Isotope: Group of words in the text with strong semantic relationships among its members - 75, 119, 129, 130, 132

Latency: The time between stimulus onset and the beginning of the ERP - 278, 284

Lexical ambiguity: Word with meanings that are highly disparate, because they often are more words that are written alike - 6, 62, 65, 66, 73, 74, 78, 82-84, 87, 95, 96, 101, 102, 194-196, 198, 199, 232, 256, 257, 260, 261, 277, 297, 322, 327, 328

Literal features: Those features in the set that are judged as descriptive, realistic or lifelike - 34, 37, 39, 43, 154, 186

Oppositions: Antonyms that are supposed to structure a text, such as 'good' vs 'evil', 'lively' vs 'dull', 'vigor' vs 'frailty'. They could easily be transformed into **factor levels** - 32-34, 119, 120, 123-125, 129, 130, 304, 306, 307, 326

Orthography: Factor that concerns the visual side of spelling (the word perceived as picture) - 61, 65, 67-69, 78, 79, 87, 233, 277, 280

Overlap: Cf. **shared set** - 16, 21-23, 30, 32, 34, 37-39, 42, 51, 54, 55, 58, 119, 127, 128, 134, 143, 144, 147, 151, 165, 168-170, 173-174, 176, 177, 203-206, 208, 209, 259, 262, 302, 307, 315, 316

Paradigm: More or less similar words of one syntactic class (e.g., nouns or verbs), which are selected to be combined in a syntagma (the sentence) - 75, 79, 119, 125-128, 131, 203, 232

Prime stimuli: Stimuli that (usually) precede the target stimuli, supposedly affecting them factorially - 11, 52, 73, 81-84, 201, 232, 233, 235, 236, 238, 240-242, 245, 282

Reaction time (RT): The amount of time needed to respond to a stimulus - 7, 26, 35, 58, 63, 78, 79, 81, 82, 85, 86, 97, 103, 197, 198, 200, 203, 209, 231-238, 241, 243-245, 247, 256, 257, 259, 260, 261, 278, 279, 282-284, 300, 327

Reading time: The amount of time needed to read a sentence - 233, 239-242, 282

Relations: Feature type that clarifies the link among other features or with the stimulus - 8, 10, 14, 17, 47-54, 58, 59, 68, 69, 76, 77, 122, 123, 130, 132, 147, 149-151, 156, 163-166, 169, 170, 172, 173, 177, 178, 197, 198, 206, 207, 210, 243, 259, 271, 273, 275, 276, 278, 324

Release time: Time between stimulus onset and leaving the home key to execute a decision - 247

Salience: The extent to which a feature is prominent or conspicuous in the feature set, expressed by a numerical weight - 51, 63, 72, 119, 135-145, 210

Semantic equality: The exact identity of meaning of two letter strings - 146, 147

Semiotics: The study of the cultural function of signs, such as symbols and icons - 4, 7, 122, 130, 132

Shared set: The number of features that two or more stimuli have in common - 14-21, 23, 25, 34, 37, 51, 52, 59, 61, 95, 119, 127, 134, 138, 139, 141-146, 148, 159-168, 171, 172, 174, 176-178, 186-188, 192-196, 198-200, 204, 205, 207, 208, 210, 212, 218, 228, 243, 244, 258, 259, 262, 275, 307, 315-317, 324, 328-330

Significance: The probability that an effect is greater than chance, *not* that it is 'highly important', 'rich in meaning' or 'meaningful' as a word in a poem can be meaningful - 87, 89, 90, 254, 320

Similarity: The resemblance between stimuli, supposedly based on a measure of their shared and distinctive features - 7, 13-18, 20, 21, 23, 29, 31, 35, 37, 39, 47, 48, 50-52, 54, 63, 68, 76-79, 85, 87, 120, 122, 124-126, 129-144, 146, 148, 186-189, 202, 203, 206, 207, 209, 225, 324, 327

Simile: A comparison with the preposition 'like', for example, 'man is like a machine' as opposed to the metaphor 'man is a machine' -

Speed-accuracy trade-off: When subjects are instructed to make rapid decisions in an RT experiment, this will be at the cost of making more errors - 254, 257

Syntagma: Sentence build up by matching words from different syntactic classes (e.g., 'child' and 'sleeps'). Cf. **paradigm** - 120, 131, 203

Target stimuli: The imperative stimuli, which are supposed to evoke the expected effect; in the present study, the *B-terms* - 3, 71, 82, 83, 245, 282

Trial: Experimental sequence of stimuli, for example, a prime followed by a target - 10, 149-151, 246, 282-284, 324

Variability: The extent to which values differ from the mean and differ among each other - 144, 190

Variance: Cf. **variability**

Samenvatting proefschrift:

Metaforen zijn dagelijks taalgebruik. "Die jongen is een ezel", "zij vocht als een leeuw" en "wat een rat is die man" worden niet letterlijk opgevat en zijn toch geen onzin. Zelfs de moeilijke metaforen in een gedicht worden begrepen ("de dood is een muur"). Welke processen benutten de hersenen om de moeilijke metaforen in een gedicht te interpreteren? Hoe zit dat met letterlijke varianten ("de dood is het einde") of onzin ("de dood is een snars")?

Allereerst heeft dit te maken met de soort associaties die zij opwekken (de figuurlijke of symbolische associaties), de hoeveelheid associaties die gelijk zijn tussen de twee termen "dood" en "muur", de snelheid waarmee ze verwerkt worden (ze tenderen langzamer te zijn dan letterlijke zinnen en onzin-zinnen) en de hoogte van de amplitude die zij opwekken in de hersenen wanneer het EEG wordt gemeten. Het blijkt namelijk dat metaforen een hoger onverwachtheidspotentiaal opwekken dan letterlijke zinnen en op andere locaties in de hersenen worden verwerkt dan onzin-zinnen, een vondst die nog nooit eerder gedaan is. Het procesmodel dat uit het onderzoek voortvloeide heeft consequenties voor de computerlinguïstiek en het onderwijs.

Persbericht:

WAT DOEN JE HERSENEN MET EEN METAFOOR?

Johan F. Hoorn

Een metafoor is een zin die je niet letterlijk kunt interpreteren, maar die ook geen onzin is. We gebruiken metaforen de hele dag: "de computer moet even nadenken", "nu loop ik over het pad met de muis" en "o, o, wat een ezel is dat ding toch". Niemand maakt er bezwaar tegen dat computers niet na kunnen denken (ze schakelen slechts), dat er helemaal niet gelopen wordt over een pad (probeer je schoenen maar eens in het beeldscherm te krijgen), dat het besturingsapparaatje geen muis en de machine geen ezel is. Het is zelfs zo dat als we alleen de technische woorden zouden gebruiken en de computer alleen in wiskunde zouden aanspreken, wij er geen barst meer van zouden begrijpen. Metaforen zijn essentieel voor menselijke communicatie en blijven niet beperkt tot het werken met computers; denk maar eens aan spreekwoorden en gezegden, gedichten en romans. Ook het onderwijs maakt er veel gebruik van: "opgelet, het atoom is een bouwsteen der materie".

Hoe doen die hersenen dat? Welke processen onderscheiden metaforen ('de man is een zwijn') van letterlijke zinnen ('de man is een mens') van onzin ('de man is een gros')? Onderzoek aan de Vrije Universiteit werd verricht naar de verschillende soorten associaties die door metaforen, letterlijken en onzin worden opgewekt bij proefpersonen, de verhouding van de *A-term* ("man") met de categorie van de *B-term* ("man" hoort niet tot de categorie "zwijnen" en "snarsen", maar wel tot de categorie "mensen"), de snelheid waarmee de diverse zinstypen verwerkt worden (reactietijden) en naar het effect van de diverse zinstypen op de hersenen, door het meten van breinpotentialen (EEG).

De resultaten waren als volgt. Woorden wekken zowel letterlijke als figuurlijke associaties op. 'Vuist', bijvoorbeeld, wekt de letterlijke associaties 'vingers, hand, kootje, nagels' op en de figuurlijke (symbolische) associaties 'woede, verzet, staking, demonstratie'. Voor letterlijke zinnen werd de connectie tussen *A*- en *B-term* (gezien als het aantal gelijke associaties tussen bijvoorbeeld "man" en "mens") alleen gevormd met letterlijke associaties. Voor metaforen was dit net zo. Echter, de connectie voor metaforen werd ook nog gelegd tussen associaties die letterlijk waren voor "man" en figuurlijk voor "zwijn". En andersom: figuurlijke associaties voor "man" waren letterlijk voor "zwijn". Voor de onzin-zinnetjes werden er nauwelijks connecties gelegd. Er waren nauwelijks associaties gelijk voor zowel "man" als "gros".

Uit de reactietijden bleek dat onzin het snelst verwerkt werd. Dat is te begrijpen als je bedenkt dat er nauwelijks enige associatie gelijk is en je dus zo klaar bent met het opbouwen van de connecties. Letterlijken en metaforen gingen even snel, hoewel metaforen ernaar negen het langzaamst verwerkt te worden. Dit is te begrijpen als je bedenkt dat letterlijke zinnen maar één soort connectie opbouwen (de gelijke letterlijke associaties) en metaforen twee soorten connecties (de gelijke letterlijke associaties én de gelijke associaties die letterlijk zijn voor het ene woord, maar figuurlijk voor het andere). Dat kost meer moeite, dus word je langzamer.

Het meten van breinpotentialen tijdens het lezen van metaforen was een novum. Nooit eerder was het gebeurd (bij Letteren noch bij Psychologie) dat er EEG werd gemeten tijdens de verwerking van metaforen. Wel was er al onderzoek gedaan naar de effecten van onzinzinnetjes, zoals "hij belegde zijn brood met boter en sokken". Het woord "sokken" liet een amplitude in het EEG zien die niet aanwezig was voor een normale afsluiting als "kaas". Dit onverwachtheidseffect werd benut om te zien of de hersenen net zoveel moeite hadden met metaforen als met onzin, of dat ze net zo makkelijk waren als letterlijke zinnen. Het bleek dat metaforen ongeveer hetzelfde onverwachtheidseffect opwekten als onzin. Letterlijke zinnen deden dat niet. Bovendien werden metaforen op iets andere locaties in de hersenen verwerkt dan onzin en letterlijken. Dit resultaat was te verklaren uit het feit dat letterlijken een *A-term* "man" benutten die behoort tot de categorie van de *B-term* ("mens"), maar dat metaforen en onzin dat niet doen ("man" behoort niet tot de "zwijnen" en "snarsen").

Het procesmodel dat hieruit voortvloeide ziet er als volgt uit. Als een zin met een *A-* en een *B-term* gelezen wordt, bekijken de hersenen eerst of de *A-term* tot de categorie van de *B-term* behoort. Dit vindt plaats gedurende de eerste 400 milliseconden nadat de *B-term* is gelezen. Als ze bij elkaar horen, wordt er geen onverwachtheidspotentiaal opgewekt. Horen ze niet bij elkaar, dan geven de hersenen een klein alarmsignaaltje af. Ze weten nu dat ze een metafoor of onzin kunnen verwachten. De volgende 1 tot 1,5 seconden worden benut om letterlijke associaties van de *A-term* te combineren met letterlijke associaties van de *B-term*, en parallel daaraan, om letterlijke met figuurlijke associaties te combineren. Wanneer dat in beide gevallen te weinig gelijke associaties oplevert, wordt de zin beoordeeld als 'onzin'. Levert dat voldoende gelijke letterlijke maar onvoldoende gelijke letterlijk-figuurlijke associaties op, dan wordt de zin beoordeeld als 'letterlijk'. Levert dat zowel veel gelijke letterlijke als veel gelijke letterlijk-figuurlijke associaties op, dan wordt de zin beoordeeld als 'metafoor'.

Wat kan zo'n procesmodel beteken voor uw computergebruik? Op dit moment kunnen computers nog niet zo goed overweg met metaforen. Ze zijn slechts in staat die commando's te interpreteren die ze al van tevoren hebben ingevoerd gekregen ("openen", "opslaan", "sluiten"). Mensen kunnen heel goed omgaan met creatief taalgebruik. Als computers weten hoe mensen dat doen (namelijk, kijk naar de categorieverwantschap van woorden en genereer zoveel mogelijk bijbehorende woorden - de associaties - om die te vergelijken op letterlijkheid en figuurlijkheid), dan zou het kunnen dat als u in de toekomst spontaan tegen uw computer zegt "wat ben je toch een ezel!", de computer antwoordt met: "neem me niet kwalijk, ik zal de fout herstellen, ik moet beter leren nadenken". De mogelijkheden voor computergebruik door leerlingen en niet-exact aangelegde mensen worden schier oneindig.

Johan Hoorn studeerde Moderne Nederlandse Letterkunde en promoveert 9 juni a.s. op een combinatie van Algemene Literatuurwetenschap en Cognitieve Psychologie.

CIP-gegevens:
Hoorn, J.F.
Metaphor and the Brain: A Behavioral and Psychophysiological Research into Literary Metaphor Processing Proefschrift Vrije Universiteit
ISBN 90-803717-1-8